



**Universidade do Minho**  
Escola de Engenharia

João Pedro Silvestre Pimenta

## **Identificação de fatores que influenciam o comportamento de peões em passadeiras**

Dissertação de Mestrado

Mestrado Integrado em Engenharia Civil

Perfil de Planeamento e Infraestruturas de Transporte

Trabalho efetuado sob a orientação da

Professora Doutora Elisabete Fraga de Freitas

Professora Doutora Jocilene Otília Costa

Fevereiro de 2018



## **AGRADECIMENTOS**

O meu reconhecimento vai para a Professora Doutora Elisabete Fraga de Freitas, sem a qual não teria sido concretizada esta dissertação. Deixo aqui um agradecimento especial a todo o seu empenho, partilha de conhecimentos e constante disponibilidade ao longo de todo o trabalho.

Para a Professora Doutora Jocilene Otilia da Costa que foi uma ajuda fulcral para a execução desta dissertação.

Para a minha família, que é imprescindível e tornou possível completar mais uma etapa importante da vida académica.

Para o Francisco Soares, pelos conhecimentos transmitidos e pela sua constante disponibilidade e vontade de ajudar ao longo de todo o trabalho.

Este trabalho enquadra-se nas atividades do projeto de investigação AnPeB – Análise do comportamento de peões com base em ambientes simulados e sua incorporação na modelação de risco (PTDC/ECM-TRA/3568/2014), foi financiado no âmbito do projeto Promover a Produção Científica e Desenvolvimento Tecnológico e a Constituição de Redes Temáticas (3599-PPCDT) e participado pelo Fundo Comunitário Europeu FEDER.



## RESUMO

A segurança rodoviária desempenha um papel crucial no dia-a-dia de qualquer ser humano. Para promover a segurança de todos os utentes das estradas é necessária uma estratégia baseada em estudos que visam o trinómio utilizador-veículo-infraestrutura. Atualmente, o foco está nos utilizadores vulneráveis com o estudo dos comportamentos e a transformação de hábitos sociais, com o propósito de diminuir a elevada sinistralidade rodoviária, e consequentemente melhorar a qualidade de vida e o bem-estar geral das populações.

No presente trabalho, pretende-se identificar os fatores que influenciam o comportamento de peões no atravessamento de passadeiras não semaforizadas e, em primeiro lugar, identificar e estabelecer um protocolo de extração desses fatores com base na análise de imagens. Para isso, foram selecionados dois locais, em Braga e Guimarães, onde foram feitas gravações de vídeo. Após a adaptação do programa de análise de imagens *Traffic Intelligence*, extraíram-se dados relativos a peões e veículos, tais como trajetórias e velocidades, a partir dos quais se estabeleceu uma base de dados.

A identificação dos fatores que influenciam o comportamento dos peões foi feita a partir da modelação da velocidade média dos peões, tendo sido desenvolvidos 5 modelos que consideram 3 momentos de um atravessamento: antes, durante, e depois do atravessamento. Um modelo contempla estes 3 momentos, outro considera os momentos antes e durante, e os restantes modelos compreendem cada momento separadamente. Na calibração dos modelos foi aplicada a técnica das Equações de Estimação Generalizadas (EEG).

Os resultados demonstraram que há um conjunto de variáveis que têm impacto na velocidade média do peão, tais como, a velocidade média do veículo, a idade do peão, entre outros. Estes resultados vão ser utilizados na parametrização e validação de cenários virtuais dedicados ao estudo do comportamento dos peões em atravessamento que no futuro serão usados para definir o risco dos peões em atravessamento.

Palavras chave: Segurança rodoviária; Comportamento de peões; Passadeira, Modelação; Velocidade média do peão.



## ABSTRACT

Road safety plays a crucial role in day-to-day life activities. To promote the safety of all road users a strategy based on studies addressing the trinomial user-vehicle-infrastructure is needed. Currently, the focus is on vulnerable users with the study of the behaviours and the transformation of social habits, with the purpose of reducing the high number of fatalities on the road and consequently improving the quality of life and the general well-being of the populations.

In the present work, it was intended to identify the factors that have influence on the behaviour of pedestrians crossing on non-signalized walkways and, firstly, to identify and establish a protocol of extraction of these factors based on image analysis. For this, two sites were selected, in Braga and Guimarães, where video recordings were made. After the adaption of the *Traffic Intelligence* image analysis program, pedestrian and vehicle data, such as trajectories and speeds, were extracted from which a database was established.

The identification of the factors influencing the behaviour of pedestrians was based on the average speed of pedestrians modelling. Five models were developed considering 3 crossing moments: before, during, and after crossing. One model contemplates these three moments, another considers the moments before and during, and the other models treat each moment separately. In the calibration of the models, the Generalized Estimation Equations (GEE) technique was applied.

The results showed that there is a set of variables that have an impact on the average pedestrian speed, such as the average vehicle speed, pedestrian age, among others. These results will be used in the parameterization and validation of virtual scenarios dedicated to the study of the behaviour of pedestrians when crossing that in the future will be used to define the risk of doing so.

**Keywords:** Road safety; Pedestrians behaviour of; crosswalk, modelling; average pedestrian speed.





# ÍNDICE GERAL

<b>1. INTRODUÇÃO .....</b>	<b>1</b>
1.1. Enquadramento.....	1
1.2. Objetivos .....	2
<b>2. ESTADO DE ARTE .....</b>	<b>3</b>
2.1. Modelos de comportamento em atravessamentos .....	4
2.1.1. Escolha discreta .....	5
2.1.2. Equações estruturais .....	7
2.2. Modelos de frequência e gravidade dos acidentes .....	8
2.2.1. Modelos de frequência dos acidentes .....	9
2.2.1.1. Regressão linear de poisson .....	9
2.2.1.2. Binomial negativa, Poisson lognormal, Poisson Inflacionado de zeros.....	10
2.2.1.3. Markov switching .....	13
2.2.1.4. Equações de estimativas generalizadas.....	14
2.2.2. Modelos de gravidade de acidentes.....	15
2.2.2.1. Multinomial logit e logit misto.....	17
2.2.2.2. Regressão ordinal .....	20
2.2.2.4. Multivariados.....	22
2.2.2.5. Hierárquicos/multinível .....	23
2.2.2.6. Redes neuronais .....	24
2.3 Considerações finais .....	25
<b>3. ESTUDO DO COMPORTAMENTO DOS PEÕES .....</b>	<b>29</b>
3.1. Secções de estudo .....	29
3.2. Definição e caracterização de variáveis .....	30
3.3. Análise de imagem recorrendo ao programa <i>Traffic Intelligence</i> .....	33
3.4. Construção da base de dados e problemas metodológicos .....	34
3.4.1 Construção da base de dados .....	34
3.4.2 Problemas na extração dos dados.....	36
3.5 Definição de um modelo de comportamento dos peões .....	38
<b>4. RESULTADOS ALCANÇADOS .....</b>	<b>39</b>
4.1. Análise estatística prévia.....	39
4.2. Correlação entre variáveis.....	43
4.3. Testes de Normalidade.....	45
4.5. Resultados dos modelos .....	48

4.5.1. Resultados modelo 1 .....	48
4.5.2. Resultados Modelo 2 .....	50
4.5.3. Resultados dos Modelo 3 a 5 .....	51
4.6. Análise de resíduos .....	51
<b>5. CONCLUSÕES E DESENVOLVIMENTOS FUTUROS .....</b>	<b>53</b>
5.1. Conclusões .....	53
5.2. Limitações e desenvolvimentos futuros.....	54
<b>6. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>57</b>

## Índice de Figuras

Figura 1.Passagem para peões, Rua Teixeira de Pascoais .....	29
Figura 2.Passagem para peões, Rua 25 de Abril .....	30
Figura 3.Definição TTC (Matsui et al., 2013).....	33
Figura 4.Programa Traffic Intelligence.....	33
Figura 5.(a) Posição do peão (3 momentos juntos); (b) Velocidade vs Tempo – 3 momentos juntos; (c) Aceleração vs Tempo – 3 momentos juntos .....	35
Figura 6.(a) Posição do veículo (3 momentos juntos); (b) Velocidade vs Tempo – 3 momentos juntos.....	35
Figura 7.União de mais que 1 objeto .....	37
Figura 8.Organigrama Modelos.....	38
Figura 9.Frequência das variáveis dummy (grupo 1): (a) Género (b) Idade (c) Movimento de cabeça.....	42
Figura 10. Frequência das variáveis dummy (grupo 2): (a) Género (b) Idade (c) Movimento de cabeça.....	42
Figura 11.Frequência das variáveis dummy (grupo 3,4 e 5): (a) Género (b) Idade (c) Movimento de cabeça .....	42
Figura 12.(a) Histograma da vmp (Grupo 1); (b) Histograma da vmp (Grupo 2); (c) Histograma da vmp (Grupo 3); (d) Histograma da vmp (Grupo 4); (e) Histograma da vmp (Grupo 5).....	47
Figura 13.Análise de resíduos Modelo 1.....	52
Figura 14.Análise de resíduos Modelo 2.....	52



## Índice Tabelas

Tabela 1 – Resumo das tipologias de modelos e critérios considerados .....	26
Tabela 2 – Estatística descritiva das variáveis do grupo 1 (3 momentos juntos) .....	39
Tabela 3 – Estatística descritiva das variáveis do grupo 2 .....	40
Tabela 4 – Estatística descritiva das variáveis do grupo 3 .....	40
Tabela 5 – Estatística descritiva das variáveis do grupo 4 .....	41
Tabela 6 – Estatística descritiva das variáveis do grupo 5 .....	41
Tabela 7 - Matriz de correlação (Grupo 1) .....	43
Tabela 8 – Matriz de correlação (Grupo 2) .....	43
Tabela 9 – Matriz de correlação (Grupo 3) .....	44
Tabela 10 – Matriz de correlação (Grupo 4) .....	44
Tabela 11 – Matriz de correlação (Grupo 5) .....	45
Tabela 12 - Teste de normalidade da vmp .....	46
Tabela 13 - Resultados GEE modelo 1 .....	48
Tabela 14 - Resultados GEE modelo 2 .....	50



# 1. INTRODUÇÃO

## 1.1. Enquadramento

A temática da segurança rodoviária tem vindo a desempenhar um papel cada vez mais expressivo no processo evolutivo da sociedade, no âmbito do planeamento urbanístico e gestão das cidades, embora em termos gerais, na Europa, o número de mortos na estrada tenha vindo nestas últimas duas décadas a reduzir significativamente (EUROSTAT, 2015). Ainda assim, o número de mortos e feridos graves continua elevado, destacando-se o referente aos utilizadores vulneráveis, particularmente o dos peões.

De acordo com o Relatório Anual – Vitimas a 30 dias da ANSR (Autoridade Nacional Segurança Rodoviária, 2015), no ano de 2015 houve 31953 acidentes com vítimas, dos quais 4668 foram acidentes quanto à sua natureza, com atropelamento de peões, correspondendo assim este valor a aproximadamente 14,60% de todos os acidentes ocorridos.

Face a estas estatísticas e à pressão política em transferir o clássico modo de transporte individual por modos suaves, esta dissertação tem como objetivo estudar fatores de risco para peões, de forma a caracterizar e observar o comportamento dos mesmos numa situação de atravessamento, para prevenir e minimizar situações conflituantes entre motoristas e peões ou outros utilizadores vulneráveis das estradas.

Atualmente existem vários estudos relacionados com a vulnerabilidade dos peões nas estradas que têm em conta fatores de natureza diversa, tais como as características físicas da estrada e o ambiente envolvente (Papadimitriou et al., 2009), a idade e o sexo (Bernhoft & Carstensen, 2008), a velocidade de circulação e condições de visibilidade (Olszewski et al., 2015), o volume de tráfego e estado de conservação das infraestruturas (Rankavat & Tiwari, 2016) e a base social e pedagógica (Al-Madani & Al-Janahi, 2006). Além disso, fatores individuais que abrangem a capacidade física do indivíduo, como reflexos ou capacidade visual (Schneider et al., 2004) também são tidos em conta. Assim, depreende-se que estes fatores e a sua interação serão determinantes na exposição ao risco de acidente num atravessamento uma vez que influenciam o comportamento adotado pelo peão durante o atravessamento.

No que se refere ao modo como esses fatores são identificados, habitualmente recorre-se a diversas técnicas de modelação. Conforme os parâmetros que estão a ser avaliados existem naturalmente técnicas mais adequadas do que outras (Papadimitriou et al., 2009). Para definir o modelo certo, é fundamental dominar e entender os vários modelos existentes para que a escolha esteja em concordância com os parâmetros que estão a ser analisados.

## **1.2. Objetivos**

O principal objetivo deste trabalho consiste na identificação de fatores que descrevam e caracterizem o comportamento dos peões no atravessamento de uma estrada. Para ser possível atingir este objetivo, previamente objetiva-se explorar as técnicas de modelação utilizadas quer em estudos de comportamento de peões em atravessamentos quer em estudos de frequência e gravidade de acidentes com peões e analisar a sua aplicabilidade às condições do estudo. Além disso, pretende-se identificar e estabelecer indicadores de comportamento que descrevam as condições de atravessamento portuguesas e por fim, ajustar um conjunto de modelos que atenderá a essas condições de atravessamento e a partir daí identificar os fatores que influenciam o comportamento dos peões.



## 2. ESTADO DE ARTE

Com o propósito de identificar as razões que levam à ocorrência de acidentes entre peões e veículos, têm surgido ao longo dos últimos anos, vários estudos neste âmbito. Isto deve-se ao crescente e rápido desenvolvimento das cidades onde se geram novas dinâmicas e consequentemente uma maior preocupação relativamente à segurança e eficiência das infraestruturas de apoio ao peão existentes.

Desta forma, foram desenvolvidos modelos com a finalidade de descrever o movimento e comportamento de peões ao longo de uma estrada, assim como modelos de frequência e gravidade de acidentes, dado que os peões são considerados os utilizadores mais vulneráveis das estradas.

Considera-se então que os modelos são ferramentas fundamentais na procura de fatores explicativos de diferentes naturezas, sendo habitualmente desenvolvidos a partir de técnicas estatísticas. Cada técnica é baseada num conjunto de pressupostos e respetivas limitações, contudo, com o progresso das mesmas nascem naturalmente novas limitações e problemas, como referenciam Savolainen, Mannering, Lord, & Quddus, (2011). Estas dificuldades estão associadas à interação de vários fatores humanos e físicos, que são muitas vezes de natureza complexa e difíceis de quantificar e, portanto, de modelar.

A modelação de peões tem problemas exclusivos que não se encontram presentes noutros modos de transporte. Tal reside no facto do movimento dos peões ser livre, o comportamento em atravessamentos não estar limitado a elementos físicos como passadeiras, e além disso há a tendência de se caminhar em grupos, ao contrário dos fluxos veiculares (Harney, 2002). A mesma fonte identifica duas principais classes de modelos, modelos macroscópicos, que examinam o sistema como um todo, e modelos microscópicos, que analisam as decisões e comportamento de cada indivíduo. Apesar desta divisão, é importante referir que algumas técnicas de modelação podem compreender a combinação de ambas as abordagens, não podendo nestes casos, serem definidos somente como macroscópicos ou microscópicos. (Harney, 2002), definiu ainda quatro abordagens diferentes na modelação macroscópica de movimentos dos peões: modelos de regressão, interações espaciais/modelos de maximização da entropia, modelos de Markov e modelos de simulação.

Papadimitriou, Yannis, & Golias, (2009), fazem uma distinção entre modelos de movimento e escolha do trajeto e modelos de comportamento num atravessamento. Nos modelos de movimento e escolha do trajeto consideram-se que existe um trajeto ótimo entre uma origem e

um destino fixo ou não. Para estes casos é geralmente usada uma modelação através de técnicas de simulação macroscópicas ou microscópicas, discretas ou contínuas, baseadas em tempos ou eventos. Já nos modelos de comportamento num atravessamento, o processo de decisão dos peões é considerado tendo em conta o tempo e/ou a localização do atravessamento. Nestes estudos é utilizada muitas vezes a *gap acceptance theory* que diz que o peão tem um intervalo de tempo crítico para realizar a passagem ou ainda a teoria da utilidade que de acordo com a qual, a utilidade de cada alternativa de travessia é um conceito latente, que é modelado como uma variável aleatória dependendo dos atributos da alternativa e as características do peão. Grande parte dos estudos baseia-se em modelos probabilísticos e determinísticos, calibrados por meio de dados observados.

O presente estudo procura o desenvolvimento de um modelo de previsão do risco para peões adaptado à realidade nacional. Desta forma, a revisão bibliográfica foca-se principalmente nos modelos que são elaborados para definir o comportamento dos peões e nos modelos que têm por base a frequência de acidentes e a gravidade dos acidentes. É importante referir que os estudos com modelos de comportamento são ainda limitados e escassos devido à complexa natureza dos fatores humanos.

## **2.1. Modelos de comportamento em atravessamentos**

Transpor o comportamento dos peões num atravessamento para modelos e análises estatísticas é algo que se tem vindo a tentar realizar de maneira a perceber como interagem os peões com a estrada, as características do tráfego e com os outros peões. Isto com o objetivo de definir medidas e adaptar cada vez mais as características da rede às necessidades do peão. Devido à natureza complexa do fator humano, a modelação do comportamento humano é ainda uma área em desenvolvimento e com muita margem de progressão (Papadimitriou, 2012).

Para descrever o comportamento dos peões foram desenvolvidos modelos, no entanto, alguns estudos observacionais não incluem necessariamente a utilização de um modelo mas seguem estruturas de análises estatísticas para comparar grupos de peões com diferentes características. O *gap acceptance* é um dos modelos mais utilizados na literatura, o qual consiste num intervalo de tempo crítico que o peão tem para realizar o atravessamento. Várias técnicas estatísticas podem ser aplicadas num modelo destes, tais como logit, probit, e derivações destes. Uma abordagem comum, ainda que mais limitada, é a do nível de serviço, que se baseia na dificuldade de realizar o atravessamento como uma medida de eficácia para o nível de serviço

do peão, ou seja, identifica as preferências no atravessamento. Há ainda os modelos de escolha discreta, que correlacionam decisões de atravessamento modeladas com base na teoria da utilidade. Uma outra perspectiva, analisa fatores psicológicos, motivacionais, de atitude e de percepção (Papadimitriou, 2012; Papadimitriou et al., 2009).

Estas abordagens incorporam elementos comportamentais de uma forma macroscópica, têm em conta o comportamento a nível local como cruzamentos ou atravessamentos urbanos. Uma avaliação do comportamento ao longo de toda a viagem seria mais vantajoso para definir estratégias de apoio ao peão, como aumentar a visibilidade das demarcações de passadeiras, instalação de “passadeiras elevadas” em zonas de maior fluxo pedonal, entre outras. Mas devido ao processo de recolha de dados e à complexidade associada à definição de uma estrutura apropriada, há ainda pouca profundidade relativamente a este tema. É também importante referir que uma outra perspectiva de análise do risco de exposição dos peões, preocupa-se mais com o movimento e escolha do trajeto dos peões. Nestas abordagens, usam-se geralmente técnicas de simulação que podem ser macroscópicas ou microscópicas e os modelos são na sua maioria estocásticos (Papadimitriou et al., 2009; Yannis et al., 2007).

### **2.1.1. Escolha discreta**

Muito utilizados em várias áreas desde a economia, psicologia, transportes, entre outras, os métodos de escolha discreta modelam a escolha de quem toma a decisão entre um conjunto de alternativas mutuamente exclusivas e coletivamente exaustivas, por outras palavras, são modelos comportamentais desagregados, tendo o peão como a unidade, projetados para prever o comportamento de indivíduos em situações de escolha de forma a obter uma representação mais realista do comportamento aquando da tomada de decisão. A probabilidade de uma pessoa escolher uma alternativa, é determinada comparando a utilidade (ou benefício ou bem-estar) de escolher essa alternativa com a utilidade de escolher outras alternativas. O modelo integra quatro fatores que são: um determinado conjunto de opções, uma lista de atributos que descrevem as alternativas, uma lista de características socioeconómicas que descrevem o decisor, e um termo aleatório que captura erros não observados. Compreende uma série de estruturas de escolha, tais como estruturas ordenadas, nested ou estruturas cruzadas, probit, entre outras (Ben-akiva & Bierlaire, 2003; Robin et al., 2009; Walker, 2001).

Modelos de utilidade aleatória, consideram que o decisor tem a capacidade de discriminar, no entanto, o analista assume que há informação incompleta e portanto, deve-se ter em conta a

incerteza, que provém de quatro fontes diferentes: erros de medição, variáveis instrumentais, características individuais não observadas, e atributos alternativos não observados. A utilidade é modelada como uma variável aleatória. As utilidades  $U$  são variáveis latentes, e as escolhas observáveis  $y$ , são manifestações das utilidades subjacentes. Assume-se que as utilidades são uma função de um conjunto de variáveis explicativas  $X$ , que descrevem o autor da tomada de decisão  $n$  e a alternativa  $i$ , isto é:

$$U_{in} = V(X_{in}; \theta) + \varepsilon_{in} \quad (\text{Eq.1})$$

Onde,  $V$  é uma função das variáveis explicativas,  $\theta$  o vetor dos parâmetros desconhecidos e  $\varepsilon_{in}$  um erro aleatório (Antonini, Bierlaire, & Weber, 2006; Papadimitriou, 2012; Papadimitriou et al., 2009; Walker, 2001; Robin, Antonini, Bierlaire, & Cruz, 2009; Robin et al., 2009)

Numa análise ao comportamento a curta distância dos peões, é proposto um modelo de escolha discreta com identificação de dois padrões principais: constrangido e sem restrições. Os constrangidos são gerados pela interação com os outros indivíduos usando um modelo leader-follower e um modelo de evitação da colisão, e os sem restrição captam os padrões comportamentais independentes das ações de outros indivíduos. Isto para obter os efeitos auto-organizadores característicos do comportamento de uma multidão. A técnica nested logit é utilizada para examinar a correlação espacial entre as alternativas, estimado pela máxima verossimilhança, que é a técnica mais utilizada para estimação destes modelos, para um conjunto de dados referentes à trajetória dos peões obtido a partir de gravações de vídeo. A inclusão de características físicas e socioeconómicas tornaria este tipo de modelos mais robusto, no entanto, iria aumentar a complexidade do modelo. Depreende-se então que a inclusão de tais características seria um dos próximos passos a considerar na modelação do comportamento dos peões de forma a obter resultados mais conclusivos (Robin et al., 2009).

Uma outra pesquisa idêntica à anterior referida, tem o objetivo de identificar as mudanças de direção e comportamento de aceleração para um modelo de peões de curto período de tempo. Seguem as duas formulações sem restrição e constrangidas com base numa estrutura de escolha discreta. Neste caso a formulação sem restrição é referente às decisões tomadas pelos indivíduos e às decisões tomadas ao nível estratégico, e a outra a acelerações constrangidas influenciadas por indivíduos percecionados como líderes. Desta forma modelos de evitação da colisão são usados para descrever os comportamentos de mudança de direção constrangidos. A correlação entre as alternativas é obtida a partir de um modelo crossed nested logit (Robin et al., 2009).

### 2.1.2. Equações estruturais

O conceito de modelos de equações estruturais compreendem diversos modelos matemáticos e métodos estatísticos que se ajustam a uma determinada rede de dados, ou seja, não se refere a uma única técnica estatística mas sim a um conjunto delas. As variáveis existentes num modelo deste tipo podem dividir-se em dois tipos: observadas e latentes. As variáveis observadas podem ser contínuas, ordinais ou nominais. Já as variáveis latentes são todas elas contínuas. Foram aplicados modelos desta tipologia para observar a relação entre o ambiente envolvente e o comportamento dos peões numa deslocação. Ainda não se sabe ao certo a força desta relação, e como tal, o uso destes modelos para compreender com mais precisão esta relação tornou-se mais frequente (Cao et al., 2007; Kline, 2010; Van Acker et al., 2007).

Desde que se começou a utilizar este modelo, o método de estimação mais ajustado é a estimativa por máxima verossimilhança. Assumindo que as variáveis observadas têm uma distribuição normal multivariada, os estimadores de máxima verossimilhança têm a capacidade de serem imparciais e consistentes (Cao et al., 2007).

Uma análise realizada na Califórnia com o intuito de investigar a relação entre as alterações efetuadas num bairro e o comportamento dos peões aquando de uma deslocação, serviu-se de um modelo de equações estruturais e usou variáveis relativas ao comportamento durante a deslocação, características do bairro, preferências do bairro, atitudes no deslocamento e sociodemográficas. O modelo foi definido da seguinte forma:

$$Y = BY + \Gamma X + \zeta \quad (\text{Eq.2})$$

Onde, Y corresponde ao vetor coluna das variáveis endógenas, X corresponde ao vetor coluna das variáveis exógenas, B corresponde à matriz dos coeficientes que representam os efeitos diretos das variáveis endógenas noutras variáveis endógenas,  $\Gamma$  corresponde à matriz dos coeficientes que representam os efeitos diretos das variáveis exógenas nas variáveis endógenas, e  $\zeta$  corresponde ao vetor coluna dos erros. Ambos os coeficiente B e  $\Gamma$  determinam a estrutura de um modelo de equações estruturais. Como resultado, uma causalidade entre o ambiente construído e o comportamento na deslocação foi observada, o que sugere que deve haver mais atenção a esta relação e como tal a criação de políticas direcionadas neste sentido (Cao et al., 2007).

## **2.2. Modelos de frequência e gravidade dos acidentes**

O ponto de partida para a modelação estatística da frequência de acidentes e respetiva gravidade é a escolha de um modelo de frequência de acidentes num determinado segmento de estrada ou intersecção, num período de tempo, seguido de uma consideração da gravidade do acidente. Ao ter em conta a gravidade do acidente, surge a necessidade de obter vários detalhes relativamente ao acidente, tais como o número de veículos envolvidos, a idade dos ocupantes, o tipo de acidente, entre outros. O uso de dados tão detalhados torna o processo cada vez mais complexo, e o cruzamento de modelos de frequência e gravidade de acidentes é imprescindível para prever e identificar fatores que afetam o nível de gravidade dos acidentes. Os fatores considerados em modelos desta génese podem servir também modelos de comportamentos, e como tal é importante conhecê-los.

Existem muitos modelos na literatura que lidam com a frequência total de acidentes. Nos seguintes capítulos irão ser abordados alguns destes modelos tais como os modelos de Poisson, binomial negativa de Poisson, Poisson log-normal, Poisson inflacionado de zeros, Markov switching.

Entre os modelos de gravidade de acidentes, várias abordagens podem ser seguidas. Multinomial logit models, ordered probit models, nested logit models, mixed logit models e ordered probit models. No entanto, a natureza dos dados destes modelos é complexa e muito detalhada, o que torna difícil previsão da gravidade dos acidentes (Anastasopoulos & Mannering, 2011).

Combinar a frequência de acidentes e a gravidade no mesmo processo de modelação cria outro tipo de complexidade porque a contagem dos acidentes para nível de gravidade não é independente. Assim, a correlação entre as contagens de acidentes com lesões específicas tem que ser considerada. Para lidar com esta correlação, métodos multivariados foram desenvolvidos, nos quais se incluem o Poisson multivariado, binomial negativa multivariado e o Poisson-lognormal multivariado. Modelos de frequência de acidentes são consideradas abordagens extremamente úteis na previsão e construção de modelos de previsão de acidentes num determinado segmento de uma estrada (Malyshkina, 2008).

## 2.2.1. Modelos de frequência dos acidentes

### 2.2.1.1. Regressão linear de poisson

Os modelos de regressão linear de Poisson têm vindo a ser utilizados como alternativa aos tradicionais modelos de regressão linear uma vez que lidam melhor com amostras de dimensões mais pequenas, além de outras características estatísticas desejadas. Definindo o domínio do tempo e espaço, e assumindo independência dos eventos, a distribuição de Poisson é plausível, ao contrário da distribuição normal que tem na sua base a regressão linear e demonstra problemas associados à não negatividade e erros de variâncias desiguais, originando problemas quanto à significância dos parâmetros. Nestes casos, um modelo de regressão de Poisson, poderá ultrapassar estes problemas (Jovanis & Chang, 1986; Jones, Janssen, & Mannering, 1991).

São muito utilizados na avaliação da frequência de acidentes ocorridos numa estrada uma vez que são uma alternativa admissível para eventos que ocorrem aleatoriamente e independentemente ao longo do tempo. Há uma grande variedade de processos relacionados com sistemas de transporte em que o resultado é representado por processos de Poisson, cuja média é alguma função com variáveis conhecidas e parâmetros desconhecidos. Este tipo de processos são denominados de regressões de Poisson e aplicam-se a falhas de equipamentos, processos industriais, modelos de migração de comportamento e frequência de acidentes rodoviários, para os quais a aptidão da técnica foi demonstrada teórica e empiricamente. Os três métodos usados para estimar os parâmetros do modelo de regressão linear de Poisson, são a máxima verossimilhança, mínimos quadrados ponderados e qui quadrado mínimo sendo que a estimação pela máxima verossimilhança é a mais conveniente segundo estudos anteriormente realizados (Jovanis & Chang, 1986; Jones et al., 1991; Lerman & Gonzalez, 1980).

Seguidamente é apresentado um exemplo de aplicação do modelo de Poisson para determinar a probabilidade de ter um número específico de acidentes na abordagem a uma intersecção  $i(n_i)$  por ano (onde  $n_i$  é um inteiro não-negativo), que é dada por:

$$P(n_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \quad (\text{Eq.3})$$

Onde,  $P(n_i)$  corresponde à probabilidade de um acidente ocorrer numa abordagem  $i$ ,  $y_i$  vezes por ano; e  $\lambda_i$  corresponde ao parâmetro de Poisson para a abordagem  $i$ , que é igual ao número

i's abordagens esperadas num ano. Os modelos de regressão de Poisson, são estimados especificando o parâmetro de Poisson  $\lambda_i$  (número esperado de acidentes por ano) como uma função das variáveis explicativas, sendo que a forma funcional mais comum é:

$$\lambda_i = \exp(\beta X_i) \quad (\text{Eq.4})$$

Onde,  $X_i$  é o vector das variáveis explicativas e  $\beta$  o vetor dos parâmetros de estimação (Poch & Fred, 1996).

É importante referir que este tipo de modelos também tem as suas limitações. Uma delas tem a ver com o facto de não terem em conta a preponderância de zero acidentes nas observações, que pode resultar em estimações divergentes dos coeficientes. São também modelos rígidos no sentido em que obrigam a que a variância dos dados seja igual à média da variável dependente  $\text{var}(Y)=\lambda=E(Y)$ . E quando tal não acontece, fenómenos de sobredispersão e subdispersão podem ocorrer (V. Shankar, Milton, & F. Mannering, 1997; Oh, Washington, & Nam, 2006).

Verifica-se sobredispersão quando variância dos dados é maior do que a indicada pelo modelos de Poisson, e subdispersão quando a variância é mais pequena do que a média obtido pelo modelo probabilístico, especialmente em casos onde a média da amostra é muito baixa (Lord et al., 2010).

Quanto maior é esta sobredispersão ou subdispersão mais ou menos subestimados são os erros padrão, e assim, o intervalo correspondente não é capaz de obter os verdadeiros valores dos parâmetros, ou seja, num estudo sobre a probabilidade de ocorrência de acidentes numa estrada, fenómenos desta natureza podem exagerar ou subestimar a probabilidade de acidentes na estrada (Miaou & Lum, 1993; Park & Lord, 2008).

Este modelo básico tornou-se um importante ponto de partida para uma família de modelos derivados da regressão linear de Poisson, os quais, vêm combater as limitações que este tem no que se refere às características dos dados.

#### **2.2.1.2. Binomial negativa, Poisson lognormal, Poisson Inflacionado de zeros**

Como já foi dito anteriormente, os modelos de regressão linear de Poisson têm a desvantagem de não lidar com fenómenos de sobredispersão, o que torna eficaz a estimação dos parâmetros muito limitada. Vários autores afirmam que uma forma de contornar este problema prende-se com a utilização de modelos de regressão binomial negativa que supera



este problema ao tolerar um valor para a variância da variável dependente, maior do que a média. Esta alternativa foi desenvolvida devido à presença deste fenómeno que é muito comum em modelos de contagem de dados e pode levar ao surgimento de erros padrão que podem ser demasiado pequenos (El-Basyouny et al., 2006)

O modelo binomial negativa deriva da (Eq.4), apresentada nos modelos de regressão linear de poisson, de tal modo que, para cada observação  $i$ :

$$\lambda_i = e^{\beta X_i + \varepsilon_i} \quad (\text{Eq.5})$$

Onde,  $\varepsilon_i$  é um termo de erro com média de 1, distribuição gamma e variância  $\alpha$ . A adição deste termo, permite que a variância seja diferente da média como se mostra em baixo:

$$VAR[y_i] = E[y_i][1 + \alpha E[y_i]] = E[y_i] + \alpha E[y_i]^2 \quad (\text{Eq.6})$$

Num estudo realizado com o objetivo de determinar a frequência de acidentes na aproximação de intersecções, a abordagem através um modelo de regressão de Poisson seria uma escolha natural, mas tendo em conta as limitações que foram anteriormente referidas, o recurso a um modelo de regressão binomial negativa, que surge como extensão do tradicional modelo de Poisson, revela-se mais adequado para estes casos uma vez que os dados demonstram características de sobredispersão (Poch & Fred, 1996; Washington, Simon P; Karlaftis, Matthew G; Mannering, 2011).

Apesar destas vantagens dos modelos de regressão binomial negativa, esta abordagem requer computações mais extensas para estimar os coeficientes do modelo e para gerar inferências estatísticas. É importante referir que a sobredispersão nestes modelos, é um indicador que o modelo não tem um bom ajuste, no entanto tem um ajuste muito melhor do que um modelo de regressão linear de Poisson (Gardner et al., 1995).

Amostras de média baixa e dimensão pequena em parâmetros dispersos fixos, apresentam-se como uma barreira nestes modelos, e neste sentido, duas importantes conclusões foram retiradas quanto ao efeito negativo que ambos os fenómenos podem originar. Primeiro, dados caracterizados por amostras de média baixa combinados com amostras de dimensão pequena, podem afetar seriamente a estimação da dispersão de parâmetros. Segundo, no caso do parâmetro de estimação ser mal estimado, as análises relativas ao mesmo podem estar erradas.

Em suma, é fundamental confirmar se todos os componentes do modelo estatístico, inclusive os parâmetros de dispersão ou o inverso, estão devidamente estimados (Lord, 2006).

Existem ainda os modelos Poisson-lognormal que foram testados por vários autores com o objetivo de encontrar uma solução mais eficaz aquando da presença de dados com características de sobredispersão. Verificou-se que são uma alternativa aos modelos de regressão binomial negativa ao apresentarem melhores resultados na presença de sobredispersão. No entanto, têm também problemas na presença de amostras de média baixa e dimensão pequena (Cameron & Trivedi, 1990).

A presença de um elevado número de zeros nos dados, levou vários autores a considerar a existência de dois estados. Tal consideração não permite uma abordagem seguida pelos modelos de Poisson ou distribuições binomiais negativas, que não têm esta capacidade, o que origina incorretas estimações dos parâmetros. Esta abordagem é utilizada em várias áreas tais como a economia, a segurança rodoviária, ciências políticas, entre outras (Lord, Washington, & Ivan, 2005; V. Shankar et al., 1997; Malyskhina & Mannering, 2010).

São modelos caracterizados por um estado duplo. No caso de serem aplicados à área da segurança rodoviária, as contagens observadas podem estar localizadas num estado perfeito, ou num estado imperfeito, isto é, são muito úteis para dados gerados por dois estados diferentes onde um é um estado de processo de contagem normal, onde as contagens podem ser inteiros não negativos gerados por um processo de contagem, e outro, um estado de contagens zero onde a probabilidade de acidentes é tão baixa que não pode ser estatisticamente distinguida de zero. Os zeros contados neste estado duplo não podem ser explicados por modelos binomiais negativo de Poisson ou modelos tradicionais de Poisson (Lord et al., 2005; Malyskhina & Mannering, 2010).

O modelo Poisson inflacionado de zeros, assume que os eventos  $Y=(Y_1,Y_2,...,Y_n)$ , são independentes e:

$$Y_i = 0 \text{ com probabilidade } p_i + (1 - p_i)e^{-\lambda_i}$$

$$Y_i = y \text{ com probabilidade } (1 - p_i)e^{-\lambda_i} \lambda_i^y / y!, y = 1, 2, \dots,$$

Onde  $y$  é o número de acidentes. Para estimar os coeficientes do modelo, o método da máxima verossimilhança pode ser aplicado, e os intervalos de confiança podem ser construídos por testes de probabilidade do rácio (J. Lee & Mannering, 2002).

Em Singapura foi analisada uma intersecção em particular para perceber quais os fatores que contribuem para o aumento de número de acidentes e definir assim as contra medidas mais eficazes aplicar. Verificou-se a presença de uma elevado número de zero acidentes que origina

muitas vezes problemas nos modelos estatísticos quando um elevado número de zero acidentes é obtido num determinado período de tempo, que pode ser devido à alta segurança do local em questão ou então porque um grande número de acidentes não foi relatado. Torna-se então fundamental tratar o modelo em dois estados diferentes e não num modelo só, porque tal abordagem poderia gerar inferências estatísticas onde um excesso de representação de zero acidentes poderia despontar. Um modelo inflacionado de zeros surge como um modelo aplicável neste tipo de situações uma vez que cria esta divisão de estados e tem um ajuste de dados mais eficaz do que os modelos de Poisson anteriormente referidos.(Kumara & Chin, 2003).

Porém, é de referenciar duas limitações associadas a este tipo de modelos. Supondo que um segmento de estrada tem zero acidentes observados num determinado intervalo de tempo, duas possibilidades devem ser consideradas. Uma é que esse segmento pode-se encontrar num estado de zero acidentes ou então, num estado de contagens normal e por acaso não houve nenhum acidente nesse período de tempo. Esta é a primeira limitação importante destes modelos, o facto de a distinção entre estas situações não ser direta. A segunda tem a ver com o facto de estes modelos não alternarem entre estados ao longo do tempo, o que pode acontecer frequentemente (Malyshkina & Mannering, 2010).

### **2.2.1.3. Markov switching**

Aplicaram-se modelos Markov switching de dois estados para analisar a frequência de acidentes, assim como a gravidade dos mesmos. Estes modelos assumem a existência de dois estados não observados na segurança rodoviária e a capacidade das entidades rodoviárias alternarem entre estes dois estados ao longo do tempo. São dois estados distintos e processados de forma diferente, ou seja, por processos de geração de dados diferentes tais como Poisson ou binomial negativa no caso de frequência de acidentes, e multinomial logit no caso de gravidade dos acidentes (Malyshkina, 2008).

Foram desenvolvidos com o propósito de contabilizar estados não observados de uma forma viável e conveniente e têm uma vasta área de aplicação noutros campos científicos. Podem ser considerados uma extensão dos modelos de Poisson inflacionado de zeros na medida em que também considera a existência dois estados diferentes, mas ao contrário destes, os modelos Markov Switching fornecem um ajuste estatístico superior e consideram o processo de comutação por segmentos de estrada entre os estados ao longo do tempo, o que permite estimar

o estado em que determinado segmento de estrada se encontra em diferentes períodos de tempo. Esta diferenciação de estados serve a necessidade de ter em conta a existência de fatores não observados ou identificados (heterogeneidade não observada), que tem influência na segurança rodoviária. Posto isto, entende-se que o um modelo Markov Switching será uma alternativa superior ao Poisson inflacionado de zeros, uma vez que contabiliza o excesso de zeros observados em dados de frequência de acidentes (Malyshkina, 2008).

A definição de um modelo Markov switching envolve o conceito de uma variável de estado não observada  $st,n$ . Tal variável determina o estado de um segmento de estrada  $n$  num período de tempo  $t$ . Por exemplo, um estudo determinou que esta variável  $st,n$  pode assumir dois valores:  $st,n = 0$  que corresponde ao estado de zero acidentes, e  $st,n = 1$  que corresponde ao estado de contagem normal ( $n=1,2,\dots,N$  e  $t = 1,1,\dots,T$ ). Seguidamente é assumido que para cada segmento de estrada, a variável de estado  $st,n$  segue um processo estacionário cadeia de Markov de dois estados no tempo. Este processo estacionário de Markov pode ser completamente especificado por probabilidades de transições independentes do tempo (Malyshkina & Mannering, 2010).

São modelos difíceis de estimar estatisticamente devido à sua complexidade e porque tem variáveis de estado não observáveis. Desta forma, não se pode usar o tradicional método de máxima verossimilhança na maioria das análises que usam modelos de Markov, mas sim uma abordagem suportada pela inferência bayesiana (Malyshkina & Mannering, 2010).

#### **2.2.1.4. Equações de estimativas generalizadas**

Introduzidos como uma extensão dos modelos dos modelos lineares generalizados, os GEE (modelos de equações de estimativas generalizadas) são um método apropriado para analisar dados discretos e correlacionados, especialmente quando estes são binários ou em formato de contagens. Não sendo modelos de regressão por si só, são utilizados como métodos para estimar modelos constituídos por dados caracterizados por correlação em série. Contabilizam a estrutura da covariância de medições repetidas, fornecendo estimações de parâmetros consistentes mesmo quando a estrutura tem erros de especificação. Modelam uma função conhecida da expectativa marginal da variável dependente como uma função linear das variáveis explicativas. Além de procurarem estimadores mais eficientes dos parâmetros de regressão, tem a grande vantagem de produzir erros padrão relativamente precisos e assim,

gerar intervalos de confiança com taxas de cobertura corretas (Abdel-aty & Abdalla, 2004; Hanley et al., 2003; Lord & Mannering, 2010; Lord & Persaud, 2000).

Com o objetivo de prever acidentes diurnos numa autoestrada conhecendo as suas características geométricas e características do tráfego em tempo real, um modelo de equações de estimativas generalizadas foi desenvolvido, uma vez que os dados demonstram características que se ajustam a este tipo de modelos. Com um número  $n_i$  de observações (caso de acidentes ou não acidentes) na localidade  $i$ , um total de localidades  $k$ ,  $y_{ij}$  denota o  $j$ -ésimo caso na localização  $i$ . Há  $\sum_{i=1}^k n_i$  casos totais. Seja o vetor dos casos, na  $i$ -ésima localidade  $Y_i = (y_{i1}, \dots, y_{in_i})'$ , e  $V_i$  a estimativa da matriz de covariância de  $y_i$ . Sendo  $x_i$  o vetor das variáveis explicativas para o  $j$ -ésimo caso na  $i$ -ésima localidade,  $X_{ij} = (x_{ij1}, \dots, x_{ijp})'$ . As equações de estimativas generalizadas para estimar  $(1 \times p)$ , onde  $p$  é o número de parâmetros de regressão e o vetor de parâmetros de regressão  $\beta$  é uma extensão da equação de estimativa de independência para dados correlacionados, são dadas por (Abdel-aty & Abdalla, 2004):

$$\sum_{i=1}^k \frac{\delta_{u_i}}{\delta \beta} V_i^{-1} [Y_i - \mu_i(\beta)] = 0 \quad (\text{Eq.7})$$

Como qualquer outro modelo, tem as suas limitações. No caso de haver valores em falta, apesar de serem estimados, a eficiência do modelo pode ficar reduzida no caso desta omissão de valores ser sistemática. E ainda, se por ventura se considerar um modelo com vários níveis de agrupamentos ou hierarquias, esta técnica não tem capacidade para lidar com eles (Hanley et al., 2003; Lord & Persaud, 2000).

### 2.2.2. Modelos de gravidade de acidentes

Presentes nesta família de modelos encontram-se os modelos logit e probit, que são abordagens muito utilizadas na análise de dados de transporte e exploram a relação entre a gravidade do acidente e os fatores intervenientes, como por exemplo, as características da estrada, o estado da estrada, as características do veículo e o ambiente envolvente. Geralmente podem-se classificar estes modelos em nominais e ordinais. Nos modelos nominais estão incluídos os multinomial logit, nested logit e logit misto. Pertencentes aos modelos ordinais, são os ordered logit, ordered probit e ordered logit misto. É de referir no entanto que existem mais modelos nestas famílias mas não são tão populares ou usados na prática. Os maiores desafios aquando da presença de dados de natureza discreta, prendem-se com a inter-relação de

dados discretos, que é um problema de seletividade, onde os dados observados são o resultado de um processo de seleção que resulta numa amostra de dados não aleatórios em categorias discretas observadas. Neste capítulo o foco principal irá incidir sobre os modelos mais usados na literatura, sendo que demonstram ser eficazes, que são o modelo multinomial logit e o ordered probit. Além destes, os modelos multivariados, modelos hierárquicos e redes neuronais também irão ser estudados, assim como o logit misto sendo que é um modelo emergente de acordo com a literatura (Washington, Simon P; Karlaftis, Matthew G; Mannering, 2011; Ye & Lord, 2014).

Variáveis contínuas diferem de proporções e probabilidades de várias formas, elas estão limitadas a zeros e uns, considerado que, em teoria, as variáveis contínuas podem assumir qualquer valor entre mais ou menos infinito. Como tal, não se pode assumir a normalidade numa proporção. Ao contrário da distribuição normal, a média e a variância na distribuição binomial não são independentes. A média é designada de  $P$  e a variância de  $P*(1-P)/n$ , onde  $n$  é o número de observações, e  $P$  a probabilidade de um evento ocorrer. Considerando os dados dispostos numa lista, em vez de tabelados, deve-se assumir que a variável tem uma média  $P$  e uma variância  $P*(1-P)$  e esta variável teria uma distribuição de Bernoulli.

Tendo uma proporção como resposta, uma transformação logística ou logit, é usada para ligar a variável dependente a um conjunto de variáveis explicativas. Esta ligação toma a seguinte forma:

$$\text{Logit}(P) = \text{Log}[P/(1 - P)] \quad (\text{Eq.8})$$

Onde, o termo entre parenteses retos refere-se à probabilidade de um evento ocorrer.

Usando a escala logit, a escala da proporção é alterada para mais e menos infinito, porque  $\text{Logit}(P)=0$ , quando  $P=0.5$ . Quando se transformam os resultados da escala logit para escala original das probabilidades, os valores esperados serão sempre entre zero e um (Elliot & Tranmer, n.d.).

A escolha entre um modelo probit ou logit não é direta, isto porque ambos os modelos têm muitas semelhanças. Na verdade, um pode ir de um conjunto de estimativas para o outro. Se um multiplica uma estimação probit por um fator, o mesmo obtém o valor aproximado correspondente à estimação logit. No entanto, existem casos onde a estimação de modelos logit e probit pode diferir substancialmente, e nesses casos é necessário uma atenção extra aquando da escolha do modelo mais ajustado. Isto acontece em casos onde há número grande de

observações com uma forte concentração de observações nas caudas da distribuição, e nestes casos, modelos logit revelam-se mais ajustados (Liao, 1994).

Considera-se um modelo probit quando se modela a probabilidade de ocorrência ou ausência de ocorrência de um evento dada uma variável independente. Considerando a ocorrência ou não ocorrência de um evento, define-se uma variável dicotômica, onde  $y=1$  se o evento ocorre e  $y=0$  se o evento não ocorre, assumindo que tais eventos e a probabilidade de eles ocorrerem dependem de uma variável independente  $x$ . Caso se pretende prever a ocorrência de um evento diretamente pela regressão de  $y$  em  $x$ , obtém-se a equação ajustada pelos quadrados mínimos para  $y$ , onde o termo de erro segue uma distribuição normal com uma média de zero (Perryman, n.d.).

Os modelos probit são muito usados na literatura para o estudo de dados com distribuição binomial. São derivados da assunção que os termos de erro das funções de utilidade têm uma distribuição normal. Captam a correlação entre todas as alternativas. Então, adotando uma notação vetorial para as funções de utilidade:

$$U_n = V_n + \varepsilon_n \quad (\text{Eq.9})$$

Onde,  $U_n, V_n$  e  $\varepsilon_n$  são vetores ( $J \times 1$ ). O vetor de termos de erro segue uma distribuição multivariada normal com uma média de zero.

$$h = F1(\mu) \quad (\text{Eq.10})$$

O inverso da função normal de distribuição cumulativa é efetivamente uma variável padronizada. Tal como o modelo logit, este modelo é usado para o estudo de resultados de variáveis binárias (Liao, 1994; Robin, Antonini, Bierlaire, & Cruz, 2009).

### **2.2.2.1. Multinomial logit e logit misto**

Modelos logit multinomiais, fornecem estimações de parâmetros consistentes na presença de uma possível subnotificação de acidentes, que é um problema recorrente nos dados relatados pela polícia, ou por outras palavras, o efeito da elasticidade nas variáveis não é afetado pela subnotificação de dados. São considerados de cálculo simples, permitem uma fácil interpretação da estimação dos parâmetros e não são muito exigentes com a qualidade dos dados (Malyskhina & Mannering, 2009; Ye & Lord, 2014; S. Jones & Hensher, 2007).

Com o objetivo de identificar os fatores que determinam a grau de lesão de acidentes com motociclistas, um modelo nominal foi aplicado onde se começou pela definição de uma função

linear que determina o resultado do grau de lesão  $i$  de um acidente com um motociclista  $n$ , como:

$$S_{in} = \beta_i X_{in} + \varepsilon_{in} \quad (\text{Eq.11})$$

Onde,  $X_{in}$  é um vetor de características mensuráveis (características dos motociclistas, características da estrada, entre outras) que determina o grau da lesão do motociclista  $n$ ,  $\beta_i$  um vetor de coeficientes estimáveis, e  $\varepsilon_{in}$  um termo de erro que têm em conta efeitos não observáveis que influenciam o grau de lesão do acidente  $n$ . Assumindo que  $\varepsilon_{in}$  tenha uma distribuição de valor extremo generalizada, o modelo multinomial logit resulta:

$$P_n(i) = \frac{e^{\beta_i X_{in}}}{\sum_{\forall I} e^{\beta_I X_{In}}} \quad (\text{Eq.12})$$

Onde,  $P_n(i)$  é a probabilidade do acidente  $n$  resultar numa lesão de gravidade  $i$  de um motociclista e  $I$  é um conjunto de resultados possíveis de gravidade de um acidente. Quanto à distribuição de valor extremo generalizada, pode também ser usada para gerar uma família de modelos que inclui o modelo nested logit (P. Savolainen & Mannering, 2007).

Distribuições de valor extremo estão classificadas em três categorias, tipo um, tipo dois e tipo três. A distribuição de valor extremo mais comum é a tipo um, sendo que tem a propriedade desejável de que os máximos de valores aleatoriamente obtidos a partir do valor extremo da distribuição tipo um, são também valores extremos da distribuição tipo um (Washington et al., 2011).

Foi conduzido um estudo no Texas, com recurso a dados de acidentes de veículos singulares envolvendo objetos fixos numa auto-estrada de dois sentidos. Verificou-se que o modelo logit misto tem uma capacidade de interpretação superior aos modelos logit multinomiais, enquanto que os modelos multinomiais têm um poder de interpretação melhor que os modelos ordered probit (Ye & Lord, 2014). Os modelos multinomiais consideram três ou mais resultados e não têm em conta a ordem explícita que pode existir nesses resultados, além disso têm uma outra limitação uma vez que são modelos suscetíveis à correlação dos efeitos não observados de um nível de gravidade de lesão para o outro, o que origina erros de especificação. Nestes casos um modelo nested logit revela-se uma formulação mais apropriada (Savolainen et al., 2011; S. Jones & Hensher, 2007).

Com o objetivo de saber quais os fatores que contribuem para a gravidade de acidentes entre peões e veículos na Coreia do Sul, um modelo multinomial logit foi calibrado para relacionar a gravidade dos acidentes com vários fatores como as condições meteorológicas, localização dos



peões, características dos peões, veículos e condutores, dispositivos de controlo de tráfego e o ambiente envolvente. Os dados foram obtidos a partir de relatórios da polícia, e os resultados obtidos irão auxiliar os engenheiros e agências controladoras de tráfego a definir medidas eficazes para promover a segurança, a redução do número de incidentes e ainda providenciar comparações para estudos futuros. Ainda neste estudo, conclui-se que muitas variáveis mostravam haver uma relação não linear com a gravidade da colisão. Por exemplo, a percentagem de condutores envolvidos em acidentes graves que estavam intoxicados foi menor do que as partes correspondentes em acidentes fatais e leves. Então, estas relações não-monotónicas reforçaram a ideia de que um modelo de resposta não ordenada revela-se mais indicado que um modelo de resposta ordenada, o que providencia suporte empírico para a escolha de um modelo multinomial logit. Em casos onde variáveis multinomiais, tais como a gravidade da lesão, são inerentemente ordenadas, o resultado é discreto, e portanto, um modelo multinomial não seria ajustado, uma vez que não tem em conta a natureza ordinal da variável dependente (Tay et al., 2011 ;Abdel-Aty, 2003).

Os modelos logit mistos pertencem à família de modelos de parâmetros aleatórios que incorporam a presença de uma possível heterogeneidade, que pode ser derivada de fatores relativos à geometria da estrada, características do tráfego e do pavimento, tipos de veículos, comportamento do condutor, entre outros (Anastasopoulos & Mannering, 2009).

Por exemplo, num caso de estudo no qual se queria verificar o efeito da flutuação do preço do combustível na quantidade de milhas percorridas por um indivíduo, é expectável que indivíduos com mais capacidade financeira se ressintam de forma diferente em relação a outros. Este efeito é acomodado num modelo de parâmetros fixos, dividindo o preço do combustível pelo salário do indivíduo como a variável independente. No entanto, podem existir outros fatores não observados que afetam a sensibilidade do indivíduo em relação à flutuação do preço do combustível, tais como, o gosto pela condução, obrigações diárias com filhos, etc. É então que um modelo de parâmetros aleatórios se revela útil, uma vez que contabiliza a heterogeneidade não observada, assumindo que os parâmetros estimados variam ao longo das observações. No entanto, quando se está a lidar com parâmetros que variam significativamente ao longo da observação, a estimação do modelo torna-se mais complexa uma vez que para cada observação, um parâmetro único é estimado para a variável em questão (Washington et al., 2011).

Como já foi referido anteriormente, o modelo multinomial logit tem várias aplicações, no entanto, tem também algumas limitações que originam estimações de parâmetros potencialmente erróneas. Um modelo logit misto têm a capacidade de ultrapassar várias fraquezas que existem nos multinomial logit ao permitir que os valores dos parâmetros variem ao longo das observações, ao contrário do multinomial logit que assume que os parâmetros são fixos ao longo das observações, ou seja, ignora a possibilidade de heterogeneidade não observada. Além disto, os seus termos de erros não são limitados a distribuições normais. Um modelo logit misto deriva da adição de um segundo termo de erro na função de gravidade, à (Eq.11) dos modelos multinomial logit:

$$S_{in} = \beta_i X_{in} + [\eta_{in} + \varepsilon_{in}] \quad (\text{Eq.13})$$

Onde,  $\eta_{in}$  é um erro aleatório com uma média de zero,  $\varepsilon_{in}$  é o termo de erro que é independente e distribuído de forma idêntica, e não depende de parâmetros subjacentes ou dados, e todas as outras variáveis são como previamente definido.

A classe de modelos logit misto assume uma distribuição geral para  $\eta_{in}$  e uma distribuição IID (erros distribuídos de forma independente e idêntica) de valor extremo tipo um para  $\varepsilon_{in}$ . Isto é,  $\eta_{in}$  pode ter várias formas de distribuição como a normal, lognormal, ou triangular. Alguns autores sugerem que existem distribuições mais apropriadas do que outras dependendo da variável em questão. Para variáveis contínuas, recomendam uma distribuição normal. Distribuições triangulares ou normais truncadas são uteis para evitar parâmetros injustificadamente grandes. (Hensher & Greene, 2003; P. T. Savolainen et al., 2011; Washington et al., 2011; J. K. Kim et al., 2010).

#### **2.2.2.2. Regressão ordinal**

Modelos de resposta ordenada reconhecem a natureza indexada das várias variáveis de resposta, tais como o ordered probit model. Ao contrário, modelos unordered multinomial ou nested logit, apesar de terem em conta a natureza categórica da variável dependente, não contabilizam a natureza ordinal as categorias de lesão. Subjacente à indexação nestes modelos, está latente o descritor contínuo de resposta. Num modelo ordered probit, o erro aleatório associado a este descritor contínuo, assume uma distribuição normal (Kockelman & Kweon, 2002; Quddus et al., 2002).

Para analisar acidentes entre veículos e peões, onde se pretendia identificar os fatores relativos à gravidade da lesão do peão numa intersecção na Flórida seguiu-se uma abordagem através de um modelo ordered probit. Conclui-se que os fatores demográficos dos condutores e peões, e a geometria da estrada, o tráfego e a envolvente, estão muito relacionadas com a frequência e gravidade dos acidentes com peões. A análise do grau de lesão de um acidente, examina a probabilidade da fatalidade ou lesão de um peão, quando estes estão envolvidos em acidente. Uma vez que se pretendia contabilizar a natureza ordinal das categorias da gravidade de uma lesão, tomou a seguinte especificação funcional:

$$Y_i^* = \beta' x_i + \varepsilon_i \quad (\text{Eq.14})$$

Onde,  $Y_i^*$  é a lesão prevista de um peão  $i$ ,  $\beta'$  um vetor de linhas de parâmetros desconhecidos,  $x_i$  um vetor de variáveis explicativas, e  $\varepsilon_i$  o termo de erro aleatório que segue uma distribuição normal. Os resultados mostraram que quanto mais idade um indivíduo tem, maior será a probabilidade do impacto causado ser severo. Também conclui-se que em áreas rurais, os acidentes tendem a ser mais gravosos que nas áreas urbanas, apesar da frequência de acidentes nessas áreas ser mais baixa. Em suma, o modelo apresenta vários fatores que contribuem para a frequência de acidentes, que vão desde características físicas, velocidade dos veículos, área de impacto e ainda percepção e reação dos condutores e peões (C. Lee & Abdel-Aty, 2005).

Para uma outra análise com o objetivo de determinar os fatores intervenientes na gravidade de uma lesão provocada por colisões entre camiões e carros, que se entende serem de natureza complexa, utilizou-se um modelo ordered probit para melhor identificar variáveis específicas que influenciem os graus de lesão em envoltimentos traseiros de dois veículos em estradas divididas. E como tal, a aplicação deste modelo revela-se apropriada uma vez que faz a análise de dados de lesão categórica ordinal. Identificam relações estatisticamente significativas entre variáveis explicativas como as características dos indivíduos e dos veículos, condições da envolvente, eventos durante um acidente e a variável dependente, que neste caso é a classe da lesão. Tem também a flexibilidade para distinguir as diferenças qualitativas entre os diferentes graus de lesão, ou seja, neste caso assume que a diferença entre uma lesão leve e sem lesão é diferente de uma lesão grave e uma morte, dada uma mudança unitária na variável explanatória (Duncan et al., 1998).

A aplicação de modelos discretos, ou modelos de gravidade de acidentes, tem também algumas limitações associadas. Uma dessas limitações está relacionada com a subnotificação de acidentes sem feridos, uma vez que são os que têm maior probabilidade de não serem

relatados às autoridades, o que pode originar estimações inconsistentes e tendenciosas dos coeficientes de um modelo ordered probability. Ao contrário, num modelo unordered multinomial logit, a estimação dos coeficientes é consistente, exceto nos termos de intercetação. O segundo problema está relacionado com restrições indesejáveis que os modelos de probabilidade ordenados colocam nas influências das variáveis explicativas.

#### **2.2.2.4. Multivariados**

Todos os modelos anteriormente comentados são uni variáveis, isto é, as contagens de acidentes de trânsito para diferentes níveis de gravidade são estimadas separadamente, o que pode produzir ineficiência nas estimativas dos parâmetros, negligenciar a informação partilhada em termos de erros não observados e tendenciosidades em bases de dados da amostra. Apesar dos dados de frequência de acidentes por gravidade serem por natureza multivariados, são muitas vezes analisados modelando cada nível de gravidade separadamente, ignorando correlações intrínsecas a cada nível gravidade. No caso de se aplicar um modelo multivariável, como o Poisson-lognormal multivariado, é possível modelar simultaneamente contagem de acidentes com a gravidade do acidente (Ma et al., 2008; Park & Lord, 2008).

Estes modelos têm a capacidade de modelar tipos específicos de contagem de acidentes, em vez de somente modelar o número total de acidentes. Sendo que para um modelo uni variável é impossível a modelação tipos específicos de contagem de acidentes, uma vez que não são independentes. E assim surgem os modelos multivariados pela sua aptidão para considerar a correlação entre os níveis de gravidade dos acidentes para cada entidade rodoviária (Lord & Mannering, 2010).

Análises com vista à segurança em autoestradas usam geralmente modelos uni variados como modelos de Poisson ou distribuições negativas binomiais para modelar contagens de acidentes para diferentes níveis de gravidade. Estas contagens são tratadas como um todo, combinando acidentes com resultados de diferentes níveis de gravidade. As consequências de tratar este tipo de problemas como um todo não são ainda bem entendidas, bem como a natureza das correlações entre níveis de gravidade e de que forma elas podem influenciar as análises de segurança.

Num estudo feito com dados relativos ao tráfego de quatro anos na Pensilvânia foram usados vários modelos estatísticos. A inclusão de modelos multivariados de efeitos aleatórios para

diferentes níveis de gravidade de acidentes demonstrou haver uma melhoria na precisão das estimações da frequência de acidentes para cada nível de gravidade e ainda um ajuste de dados significativamente melhor. O desvio padrão médio da estimação de frequência de acidentes foi reduzido em 20% o desvio padrão dos resultados mais graves foi reduzido ainda mais (41% para as fatalidades e 48% para as lesões mais graves). Em termos de classificação, verificou-se que, para os segmentos mais bem classificados, as estimativas multivariadas do custo total de colisão e do custo de colapso em excesso são consistentemente maiores do que as estimativas independentes. Estes são alguns dos efeitos positivos gerados por seguir-se uma abordagem multivariada comparativamente a abordagens uni variadas (Aguero-Valverde & Jovanis, 2009).

### **2.2.2.5. Hierárquicos/multinível**

Estruturas hierárquicas são basicamente descrições estatísticas de uma disposição de dados que são caracterizadas por respostas correlacionadas dentro de aglomerados hierárquicos. Na verdade, justificam-se pela presença de aglomerados hierárquicos, de outra forma, métodos não hierárquicos serão os apropriados.

Foi realizada uma análise em intersecções rurais com uma estrutura de dados hierárquica disposta da seguinte forma: os acidentes representam o nível mais baixo da hierarquia, enquanto a intersecção na qual o acidente ocorreu representa o nível mais alto da hierarquia. Presumindo que fatores não observados afetam a probabilidade da tipologia do acidente em intersecções, como por exemplo, pavimento em mau estado ou refletividade fraca dos sinais de trânsito, entre outros, deduz-se que uma correlação entre estes fatores não observados, a frequência dos acidentes e a tipologia dos mesmos num determinado local, pode existir. A correlação dentro de aglomerados viola a suposição de independência residual assumida em muitos métodos estatísticos. Se esta correlação não é verificada, a consequência geralmente é a atenuação dos efeitos, estimação de parâmetros parciais, erros padrão tendenciosos, ou heterogeneidade da regressão. Problemas desta génese podem ser superados através da aplicação de modelos multinível (D. G. Kim et al., 2007).

Consideram-se modelos multinível os mais indicados para casos onde os dados demonstram uma estrutura hierárquica: aglomerados, acidentes ou carros por exemplo, e um certo número de subaglomerados por aglomerado, como ocupantes dentro do carro e carros dentro do acidente. As suas aplicações iniciais foram na modelação de respostas lineares, no entanto também permitem a modelação de problemas não lineares. Identificam e quantificam a relação

dos fatores de risco, por exemplo, com a gravidade, ao mesmo tempo em que se modela a estrutura hierárquica dos dados correlacionados. São muito utilizados em dados com um número limitado de aglomerados mas um grande número de subaglomerados por aglomerado, e nestes casos, problemas de estimação de parâmetros podem surgir, assim como na qualidade do resultado das estimações. Tais problemas despontam consoante a natureza do resultado, se a distribuição é normal ou não. Como tal, as maiores dificuldades nos modelos multinível com uma estrutura hierárquica fraca, aparecem quando há um número pequeno de aglomerados (número de acidentes). Com dados de acidentes rodoviários, é possível ter um número grande de aglomerados. No entanto, baixos números subaglomerados por aglomerado podem originar problemas de convergência e estimações tendenciosas. Ignorar a possível existência de uma estrutura hierárquica nos dados é algo recorrente, e tal desconsideração pode levar estimações incertas, erros padrão incorretos e limites de confiança. Aquando da estimação de parâmetros nos modelos multinível, vários métodos podem ser utilizados. Consideram-se os tratamentos totalmente bayesianos computacionalmente viáveis, sendo que têm a vantagem de em amostras de dimensão pequena, contabilizar a incerteza associada às estimações dos parâmetros de maior nível, e providenciam medidas de incerteza exatas (Lenguerrand, Martin, & Laumon, 2006; Lenguerrand et al., 2006).

#### **2.2.2.6. Redes neuronais**

Mais recentemente, alguns autores têm vindo a usar redes neuronais artificiais na análise de gravidade de acidentes. Estes modelos procuram algum tipo de padrão nos dados e permitem potenciais relações não lineares entre níveis de gravidade de acidentes e covariáveis. Fornecem técnicas analíticas que utilizam informação de observações anteriores de modo a prever novas observações. O modo de funcionamento destes modelos é altamente complexo uma vez que foram inspirados a partir do funcionamento do cérebro humano. São compostos por processadores simples designados de nós, e números conexões entre eles. Consistem numa camada de entrada, onde os dados são apresentados à rede, e uma camada de saída que liga a resposta da rede a uma determinada entrada. Muitas vezes, para captar as potenciais relações não lineares existentes nos dados, são introduzidas camadas intermédias, designadas de camadas escondidas na arquitetura da rede. Então, as redes neuronais, “aprendem” através da adaptação do peso das suas conexões, enquanto que os nós escondidos organizam-se de forma a que os diferentes nós consigam reconhecer diferentes características do espaço total de entrada

(Delen et al., 2006; P. Savolainen & Mannering, 2007; P. T. Savolainen et al., 2011; Washington et al., 2011).

São vários os fatores que afetam o risco da gravidade de uma lesão em acidentes, tais como, a idade do ocupantes, o gênero, o uso do cinto de segurança, fatores ambientais, condições da estrada, características técnicas do veículo etc. Foi realizado um estudo com o objetivo identificar quais destes fatores influenciam a probabilidade do aumento da gravidade da lesão durante um acidente. A tipologia dos acidentes em questão inclui acidentes de colisão multi-veículo, colisões entre um único veículo e um objeto fixo e acidentes sem colisão. Problemas desta natureza foram tratados geralmente com modelos lineares onde a relação funcional entre a gravidade do acidente e fatores relacionados com a colisão era assumida como linear. No entanto, estes modelos lineares demonstram algumas limitações relacionadas com o uso de distribuições não homogêneas e têm propensão a problemas estatísticos quando a correlação entre as variáveis independentes é maior do que os níveis aceitáveis, originando modelos com erros superiores ao desejável. Sendo que as redes neuronais artificiais são capazes de captar relações não lineares entre as variáveis de previsão e a variável alvo, modelos de redes neuronais foram utilizados para estimar os efeitos de fatores significantes da colisão nos níveis de gravidade da lesão sustentado pelo condutor (Delen et al., 2006; P. T. Savolainen et al., 2011).

## **2.3 Considerações finais**

Em suma, verifica-se que a modelação de um conjunto de dados implica o conhecimento das técnicas existentes para que se obtenha resultados significativos. Há uma vasta gama de técnicas de modelação que podem ser usadas no âmbito da segurança rodoviária. O estudo destas técnicas é fundamental para a seleção do melhor modelo a ser usado na definição do conjunto de variáveis que irão integrar este estudo, isto é para a identificação de fatores que intervêm no comportamento de um peão durante um atravessamento e na possibilidade de ocorrência de acidente.

A Tabela 1 resume de uma forma muito generalizada os critérios que foram usadas por vários autores para o estudo do comportamento dos peões, frequência e gravidade de acidentes.

Como se pode verificar, para os modelos de comportamento de peões em atravessamento, há 2 critérios considerados relativos às trajetórias e acelerações de peões, que são utilizados

neste trabalho, ou seja, são critérios que, juntamente com outros, têm o potencial de descrever e identificar fatores que descrevam o comportamento dos peões, de acordo com a literatura.

Além destes, são considerados critérios relacionadas com o ambiente envolvente, onde se podem incluir larguras de passeios ou vias, estados de conservação dos pavimentos, geometrias da estrada, condições climáticas, entre outras.

Tabela 1 – Resumo das tipologias de modelos e critérios considerados

<b>Tipologia de Modelos</b>	<b>Crítérios considerados</b>
<b>Modelos de comportamento em atravessamentos</b>	<ul style="list-style-type: none"> <li>• Trajetórias de peões</li> <li>• Acelerações de peões</li> <li>• Variáveis comportamentais</li> <li>• Variáveis demográficas</li> <li>• Ambiente construído</li> </ul>
<b>Modelos de frequência dos acidentes</b>	<ul style="list-style-type: none"> <li>• Caraterísticas do veículo</li> <li>• Caraterísticas do condutor</li> <li>• Caraterísticas de tráfego</li> <li>• Caraterísticas da estrada</li> <li>• Condições climáticas</li> <li>• Número de acidentes</li> </ul>
<b>Modelos de gravidade dos acidentes</b>	<ul style="list-style-type: none"> <li>• Caraterísticas da estrada</li> <li>• Condições meteorológicas</li> <li>• Ambiente envolvente</li> <li>• Geometria da estrada</li> <li>• Caraterísticas de tráfego</li> <li>• Frequência dos acidentes</li> <li>• Níveis de gravidade</li> <li>• Tipologia dos acidentes</li> <li>• Variáveis demográficas</li> </ul>

Por último, e geralmente mais utilizadas em estudos voltados para a frequência e gravidade de acidentes, consideram-se o número de acidentes, tipologia dos mesmos, níveis de gravidade de um acidente etc. Tendo tudo isto em conta, a definição das variáveis que irão integrar o presente estudo apoiam-se nestas considerações para que se possam identificar quais as que realmente intervêm no comportamento dos peões.







### 3. ESTUDO DO COMPORTAMENTO DOS PEÕES

Neste capítulo pretende-se apresentar e explicar passo a passo a metodologia adotada neste estudo, desde a escolha das secções de estudo até à formulação de modelos que caracterizem o comportamento dos peões.

O estudo do comportamento dos peões divide-se em 5 fases fundamentais:

- 3.1. Seleção das secções de estudo;
- 3.2. Definição e caracterização de variáveis;
- 3.3. Análise de imagem e extração de dados;
- 3.4. Construção da base de dados e problemas metodológicos;
- 3.5. Modelação do comportamento dos peões.

#### 3.1. Secções de estudo

As secções de estudo escolhidas para a análise, foram 2 passagens para peões, junto a intersecções, sendo que uma localiza-se em Guimarães na rua Teixeira de Pascoais e a outra na rua 25 de Abril em Braga (Figura 1 e 2). Para ambas, recorreu-se a gravações de vídeo de aproximadamente 35 minutos por questões de logística com início às 17 horas da tarde, sendo que é uma hora de ponta como se pode constatar em contagens de tráfego anteriormente realizadas, e por isso há um grande volume de veículos e peões nas estradas.



Figura 1. Passagem para peões, Rua Teixeira de Pascoais



Figura 2. Passagem para peões, Rua 25 de Abril

A rua Teixeira de Pascoais (Figura 1) contém uma faixa de rodagem de aproximadamente 7 metros com 2 vias (uma em cada sentido). O volume de ponta horário pedonal é de 740 peões/hora e o rodoviário é de 1272 veículos/hora.

No caso da rua 25 de Abril (Figura 2), a passagem para peões situa-se numa estrada com uma faixa de rodagem de aproximadamente 7 metros de largura e 2 vias com apenas um sentido de tráfego. O volume de ponta horário pedonal é de 779 peões/hora e o rodoviário é de 790 veículos/hora.

Pode-se considerar que são ruas com características idênticas, tanto físicas como funcionais.

### 3.2. Definição e caracterização de variáveis

A definição de variáveis para a elaboração da base de dados apoiou-se na revisão bibliográfica, incluindo nomeadamente, variáveis demográficas e comportamentais que são utilizadas para a descrever o comportamento dos peões. No que se refere ao ambiente rodoviário, como o estudo está limitado a apenas duas secções de estudo, não há variabilidade suficiente pelo que não se consideram estas variáveis. Considera-se ainda as características dos veículos e suas relações.

Os dados incluídos na construção da base de dados referem-se então à extração de informação relativa a peões e veículos. No que se refere aos peões, o conjunto de variáveis é o seguinte:

- Género: Identificação do sexo do peão
- Idade: Identificação do grupo etário a qual o peão pertence com base em 2 grupos previamente definidos como idoso (mais de 65 anos) e não idoso (menos de 65 anos).
- Veículo em aproximação pela esquerda ou direita do peão: Identificação do veículo que se encontra à esquerda ou direita do peão no momento em que este efetua o atravessamento, caso exista.
- Número do Frame, Posição (x e y), Velocidade (x e y), norma: Nestas variáveis irão entrar todos os frames do vídeo em que o peão existe. E relativamente a cada frame, este terá uma posição em X e Y, e uma velocidade em X e Y bem como a norma da velocidade. Todos estes valores são fornecidos pelo programa utilizado.
- Velocidade Instantânea em cada frame (vip), velocidade média 2 segundos antes do atravessamento (vmp) , velocidade média durante o atravessamento (vmp) , velocidade média 2 segundos depois do atravessamento (vmp) : A atribuição dos valores destas variáveis a todos os peões, nem sempre é possível visto que há casos em que não é possível obter a informação total relativamente ao peão. Posteriormente será explicado como é que isto acontece.
- Aceleração Instantânea em cada frame, Aceleração média 2 segundos antes do atravessamento (ac), Aceleração média durante o atravessamento (ac), Aceleração média 2 segundos depois do atravessamento (ac) : A atribuição dos valores destas variáveis a todos os peões, nem sempre é possível visto que há casos em que não é possível obter a informação total relativamente ao peão. Posteriormente será explicado como é que isto acontece.
- Frame de início e fim de atravessamento: Regista-se o frame em que o peão chega ao início do atravessamento, e o frame em que este termina o atravessamento.
- Frame de início de travessia: Frame correspondente ao momento em que peão inicia o atravessamento.
- Sentido: Esta variável indica se o peão faz a travessia no sentido 0 ou sentido 1. Tais sentidos são previamente definidos.
- Movimento de cabeça (mc) : É atribuído o valor 0 caso o peão não olhe no momento em que vai iniciar o atravessamento, e o valor 1 caso o peão olhe.
- Grupo ou não: Atribui-se o valor 0 caso o peão faça o atravessamento sozinho, e o valor 1 caso este vá em grupo.

- CPA: Que por definição se designa por “*Closest-Point-Of-ApproachJoin*”, corresponde à distância direta mais próxima possível entre 2 objetos dinâmicos. Neste trabalho, a variável que foi chamada de CPA é idêntica à sua definição de origem, mas a distância é calculada num determinada instante de tempo, e não ao longo do tempo. Esta distância é determinada a partir das coordenadas de ambos os objetos, extraídas a partir do programa utilizado

No caso dos veículos:

- Número do Frame, Posição (x e y), Velocidade (x e y), norma: Nestas variáveis irão entrar todos os frames do vídeo em que o veículo existe. E relativamente a cada frame, este terá uma posição em X e Y, e uma velocidade em X e Y bem como a norma da velocidade. Todos estes valores são fornecidos pelo programa utilizado.
- Velocidade Instantânea em cada frame (viv) , velocidade média 2 segundos antes do atravessamento (vmv), velocidade média durante o atravessamento (vmv), velocidade média 2 segundos depois do atravessamento (vmv).
- Aceleração Instantânea em cada frame, Aceleração média 2 segundos antes do atravessamento (acv), Aceleração média durante o atravessamento (acv), Aceleração média 2 segundos depois do atravessamento (acv).
- Sentido: Esta variável indica se o veículo faz a travessia no sentido 0 ou sentido 1. Tais sentidos são previamente definidos.
- Chegada à passadeira: Frame correspondente ao momento em que o veículo chega à passadeira.
- Partida da passadeira: Frame correspondente ao momento em que o veículo parte da passadeira.
- TTC: Variável denominada de “time to collision”, em português, tempo para colisão. Consiste no tempo entre um ponto de aproximação e a colisão potencial entre um peão e um veículo.

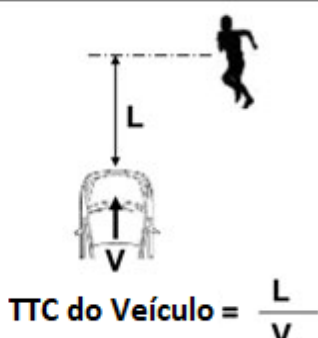
<b>Objeto</b>	<b>Veículo</b>
<b>Tempo</b>	<b>TTC (time-to-collision)</b>
<b>Definição</b>	 $\text{TTC do Veículo} = \frac{L}{V}$

Figura 3.Definição TTC (Matsui et al., 2013)

### 3.3. Análise de imagem recorrendo ao programa *Traffic Intelligence*

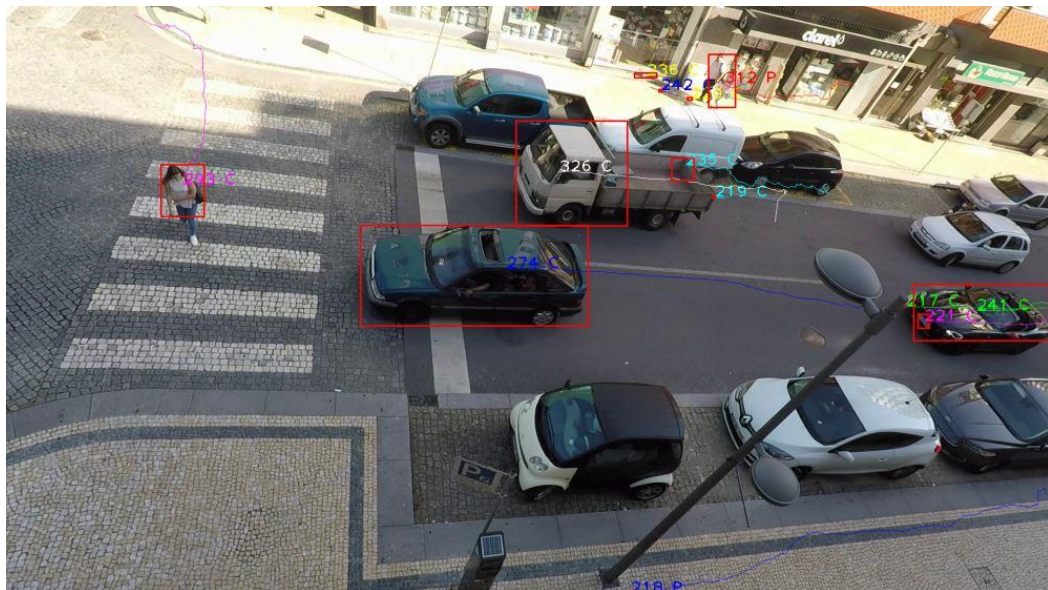


Figura 4.Programa Traffic Intelligence

Para a análises dos vídeos foi usado o programa *Traffic Intelligence* que foi desenvolvido no âmbito de um projeto Europeu (Jackson et al., 2013). Este programa é de acesso livre, contudo com informação sobre o seu funcionamento muito escassa. Além disso, na última versão disponibilizada, a maioria dos módulos não estão acabados, o que restringe as funcionalidades do programa àquilo que aqui foi usado. Após algum trabalho de adaptação do programa, foi

possível identificar e diferenciar peões e veículos, desenhando caixas em torno dos objetos, chamadas de “bounding boxes” (Figura 4), e ainda retirar informação relativa aos mesmos, tal como velocidades e trajetórias, que deram origem a outras, já referidas.

Previamente é necessário fazer uma calibração dos parâmetros do programa para cada vídeo, para que o ajuste seja o mais preciso possível, ou seja, para que deteção de veículos e peões tenha o menor erro possível. A calibração consiste em ajustar uma série de parâmetros relacionados com a qualidade de deteção de “features” e com as distâncias entre “features” que permitirão o agrupamento das mesmas, tornando capaz a assimilação e identificação dos vários objetos por parte do software. Este processo leva algum tempo e mostra-se algo árduo porque se baseia num processo de tentativa erro.

A extração dos dados é feita automaticamente, porém não dispensa verificação das trajetórias dos objetos identificados (peões e veículos).

Devido às limitações do programa, os dados de veículos e peões não podem ser todos integrados na base de dados porque as trajetórias não foram identificadas na totalidade por motivos que serão mencionados no subcapítulo 3.4.2.

### **3.4. Construção da base de dados e problemas metodológicos**

#### **3.4.1 Construção da base de dados**

A construção da base de dados começou pela separação da informação extraída pelo programa por peões e veículos. No caso dos peões, o próprio programa atribuiu um número a cada um e regista a trajetória e velocidade respetivas ao longo do tempo. A partir daqui construíram-se gráficos de posição bidimensionais, gráficos de velocidade e aceleração em função do tempo (Figura 5). Nestes gráficos foram distinguidos 3 momentos: antes do atravessamento, durante o atravessamento e depois do atravessamento.



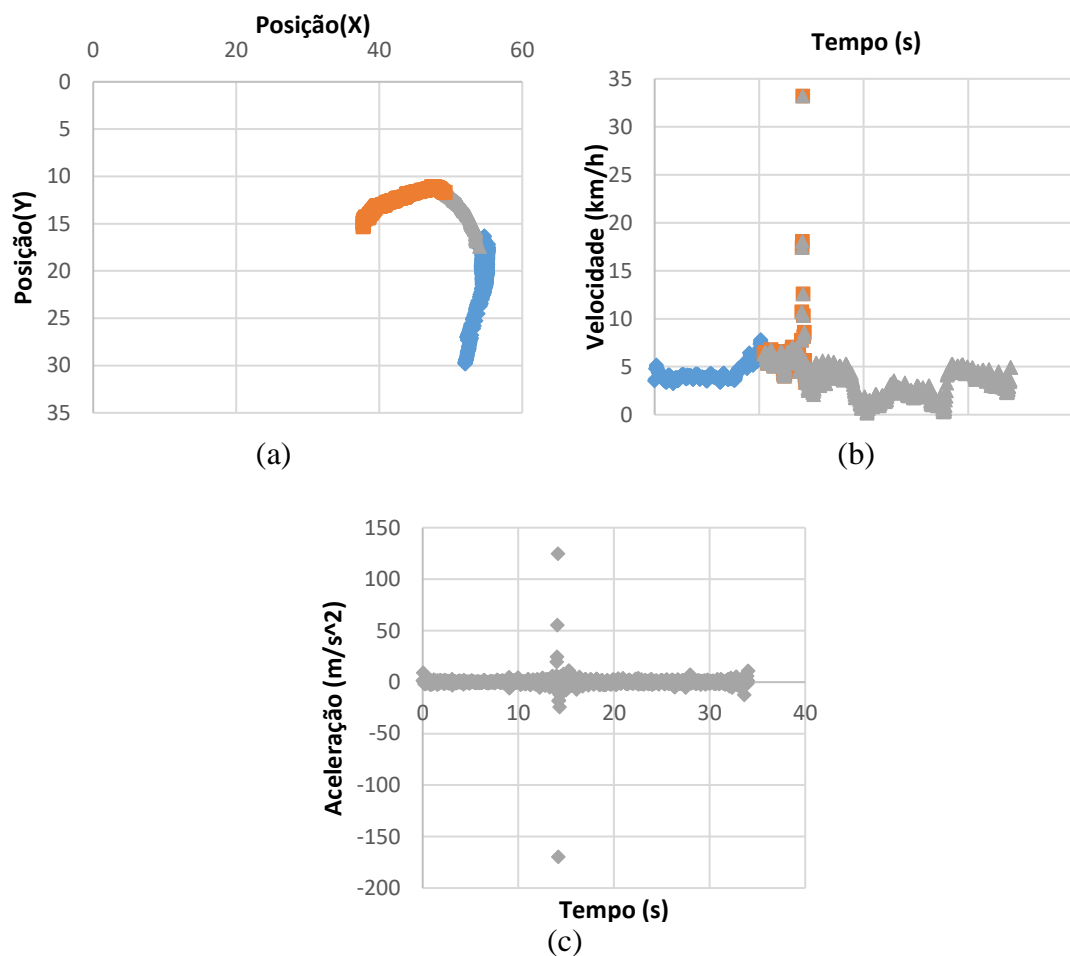


Figura 5.(a) Posição do peão (3 momentos juntos); (b) Velocidade vs Tempo – 3 momentos juntos; (c) Aceleração vs Tempo – 3 momentos juntos

O peão que aqui é usado como exemplo, tem aquando da sua travessia, a passagem de um veículo. Os gráficos desse veículo são apresentados na Figura 6. Para os veículos, apenas gráficos de posição e velocidade foram elaborados.

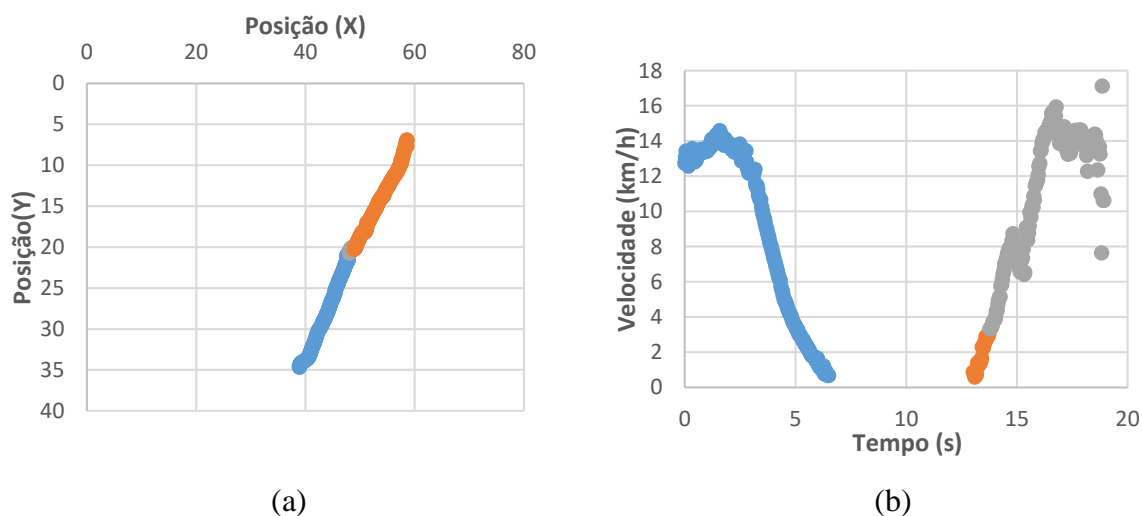


Figura 6.(a) Posição do veículo (3 momentos juntos); (b) Velocidade vs Tempo – 3 momentos juntos

### 3.4.2 Problemas na extração dos dados

Como é possível ver na Figura 6, no gráfico de velocidade-tempo, há um intervalo de tempo sem informação. Tal acontece porque o veículo para, para dar passagem ao peão, o que implica manualmente juntar a informação antes e depois da paragem sendo que o programa atribui ao mesmo objeto uma identificação diferente. Seguidamente, são apresentados problemas que surgiram durante a extração de dados:

- A gravação vídeo de um peão que vai efetuar um atravessamento tem a sua trajetória definida pelo programa, que pode ser afetada por sombras, árvores que bloqueiem o peão na imagem, a passagem de um autocarro que esconde o peão da imagem ou simplesmente porque não foi possível a extração da informação na íntegra. Ou ainda, o caso em que peão é apenas detetado pelo programa a meio do atravessamento, e nesse caso não vai existir informação relativa ao momento antes do atravessamento. Situações deste género limitam a extensão da base de dados, uma vez que podem impossibilitar a extração de informação relativa aos 3 momentos do atravessamento.
- Erros devido à calibração do programa, podem originar uma situação em que peões que estão bastante distanciados, são detetados como um só e criada uma caixa em torno deles, impossibilitando a contabilização de grande parte dos peões visíveis naquele momento. O mesmo pode ocorrer entre veículos. Este tipo de erro, surge com alguma frequência por questões relacionadas não só com a calibração do programa, mas também com o posicionamento da câmara.
- Uma outra situação muito recorrente neste programa, tem a ver com a proximidade de objetos, isto é, quando os peões estão em grupo (mais que 1 peão), não é possível identificar cada um deles separadamente. A informação relativa a 2 ou mais peões que efetuam o atravessamento em grupo, é gerada como se fosse apenas 1 objeto, como se pode verificar na seguinte Figura 7.

Como é possível ver na Figura 7, há 2 grupos de peões distintos. Em ambos os casos, o programa deteta e cria uma caixa em volta desse mesmo grupo, de forma a definir a sua velocidade e trajetória. Torna-se impossível individualizar a informação para cada um dos peões.

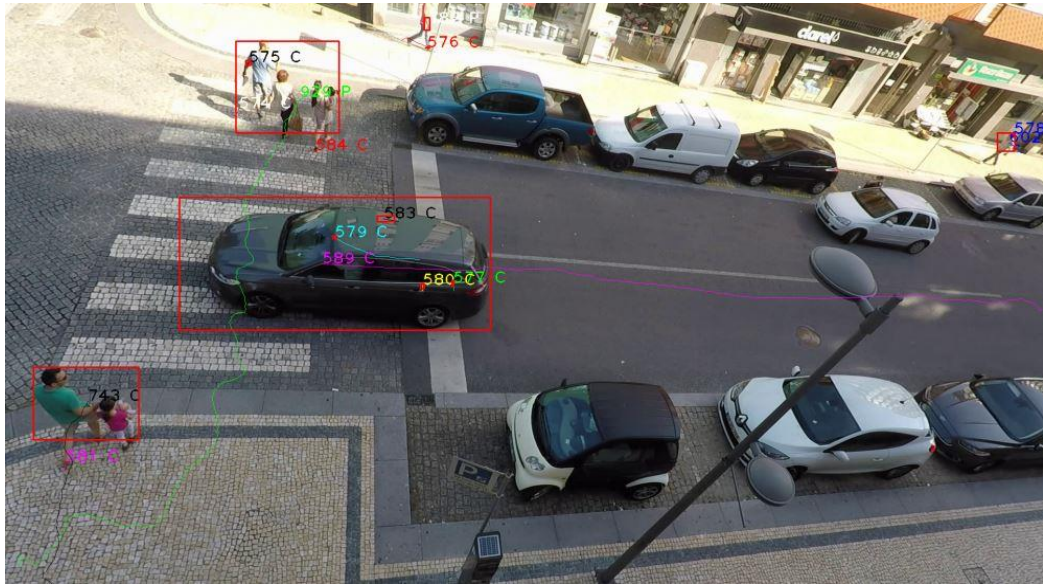


Figura 7. União de mais que 1 objeto

- Também acontece com alguma frequência, o objeto ser detetado e seguidamente parar (como por exemplo um veículo que para na passadeira, para um peão iniciar o atravessamento). O veículo que já tinha sido detetado, perde a identificação lhe tinha sido atribuída, e até ao momento em que volta a iniciar a marcha, deixa de haver informação e é lhe atribuída uma nova identificação, perdendo-se assim a informação nesse espaço de tempo. Em termos gráficos, pode verificar-se esta situação na Figura 6, onde é notória a falta de informação durante um determinado espaço de tempo.

Esta tipologia de problemas, gerou uma quantidade considerável de valores omissos. Tal deve ser tido em consideração aquando da escolha do método mais adequado para aplicar aos modelos que irão ser definidos.

### 3.5 Definição de um modelo de comportamento dos peões

É importante nesta fase do projeto elaborar um modelo que represente a realidade da melhor forma possível. Como já foi referido, a análise foi dividida em 3 momentos diferentes, a partir destes 3 momentos foram definidos 5 grupos de variáveis. Para cada grupo de variáveis será desenvolvido um modelo que relaciona a vmp com os fatores explicativos já apresentados.

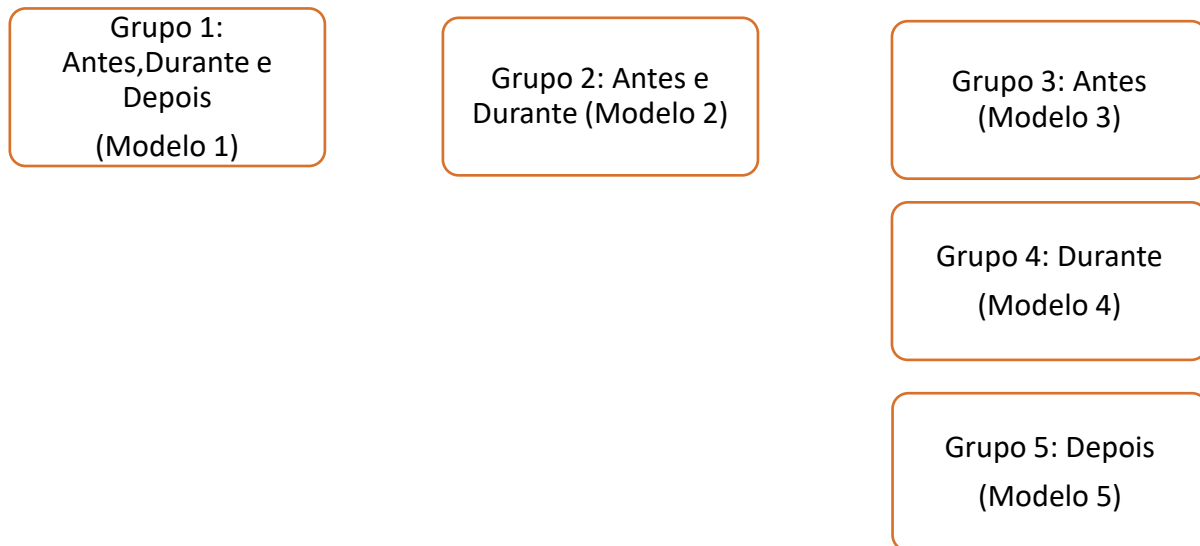


Figura 8. Organograma Modelos

Definidos os grupos de variáveis, parte-se para descrição estatística dos dados, onde será feito um estudo das variáveis definidas quanto às suas características, correlação entre variáveis e testes de normalidade. Depois deste estudo, o próximo passo será a aplicação das técnicas estatísticas de equações de estimativas generalizadas (GEE) a todos os grupos de variáveis. Os resultados são apresentados no capítulo seguinte.

## 4. RESULTADOS ALCANÇADOS

No presente estudo foram elaborados 5 modelos para avaliar o comportamento dos peões com o programa SAS® (Statistical Analysis System). Cada modelo compreende 5 grupos de variáveis. Inicialmente, para cada grupo de dados obteve-se uma análise estatística passando pelas estatísticas descritivas para sumarizar os dados, a construção de gráficos de frequência de variáveis, determinação de matrizes de correlação e testes de normalidade. Concluída esta primeira etapa foram ajustados os modelos e feita a interpretação dos resultados.

### 4.1. Análise estatística prévia

A primeira etapa da análise dos resultados inicia-se com a estatística descritiva, com o objetivo de descrever e resumir os dados. Referente ao Grupo 1 (agregado) pode observar-se a Tabela 2 com a descrição estatística de cada uma das variáveis definidas, a Tabela 2 relativa ao Grupo 2, e ainda as Tabelas 4,5 e 6 referentes aos restantes grupos de variáveis.

Tabela 2 – Estatística descritiva das variáveis do grupo 1 (3 momentos juntos)

Variável	N	Média	Desvio padrão	Somatório	Máximo
ac	189	-0.0359	0.1853	-6.7748	0.3431
acv	132	-0.0764	2.0328	-10.0881	6.4819
viv	161	18.0779	11.2216	2911	59.2198
vip	188	4.5486	2.2697	855.1539	25.6544
vmp	189	4.2235	1.0133	798.2394	9.5573
vmv	131	17.4561	9.2871	2287	39.3435
cpa	162	12.9194	7.6008	2093	42.9279
ttc	164	38418	48786	6300566	100000

Tabela 3 – Estatística descritiva das variáveis do grupo 2

<b>Variável</b>	<b>N</b>	<b>Média</b>	<b>Desvio padrão</b>	<b>Mínimo</b>	<b>Máximo</b>
<b>ac</b>	126	0.0033	0.1264	-0.3846	0.2608
<b>acv</b>	102	-0.4972	1.9407	-12.2888	3.7723
<b>viv</b>	101	17.5402	10.4747	1.4580	43.4772
<b>vip</b>	126	4.2915	1.8549	1.2624	15.3954
<b>vmp</b>	126	4.1499	1.0796	1.5034	9.5573
<b>vmv</b>	101	18.6764	9.7879	0.8795	39.3435
<b>cpa</b>	102	15.7149	7.6175	3.3115	42.9279
<b>ttc</b>	101	5.6019	3.7749	0.8688	31.0477

Tabela 4 – Estatística descritiva das variáveis do grupo 3

<b>Variável</b>	<b>N</b>	<b>Média</b>	<b>Desvio padrão</b>	<b>Mínimo</b>	<b>Máximo</b>
<b>ac</b>	63	-0.0217	0.1601	-0.3846	0.2499
<b>acv</b>	39	-1.0493	2.833	-12.289	1.3556
<b>viv</b>	39	19.0962	10.2404	2.1382	36.3443
<b>vip</b>	63	4.4565	2.2402	1.3322	15.3954
<b>vmp</b>	63	3.8362	1.0629	1.5034	6.2224
<b>vmv</b>	38	18.7155	10.5522	1.2565	38.9193
<b>cpa</b>	39	18.5941	7.9344	3.3115	42.9279
<b>ttc</b>	38	6.0596	2.7813	0.8688	13.8084

Tabela 5 – Estatística descritiva das variáveis do grupo 4

Variável	N	Média	Desvio padrão	Mínimo	Máximo
<b>ac</b>	63	0.0282	0.0728	-0.1800	0.2608
<b>acv</b>	63	-0.1555	0.9517	-4.5106	3.7723
<b>viv</b>	62	16.5614	10.5840	1.4580	43.4772
<b>vip</b>	63	4.1265	1.3649	1.2624	7.2943
<b>vmp</b>	63	4.4635	1.0101	2.9318	9.5572
<b>vmv</b>	63	18.6528	9.3845	0.8795	39.3435
<b>cpa</b>	63	13.9324	6.8919	4.3758	28.2114
<b>ttc</b>	63	5.3258	4.2616	1.8295	31.0477

Tabela 6 – Estatística descritiva das variáveis do grupo 5

Variável	N	Média	Desvio padrão	Mínimo	Máximo
<b>ac</b>	63	-0,1141	0,2502	-1,0862	0,3431
<b>acv</b>	30	1,3544	1,6757	-1,3419	6,4820
<b>viv</b>	60	18,9832	12,4175	1,5176	59,2198
<b>vip</b>	62	5,0714	2,8857	2,5978	25,6544
<b>vmp</b>	63	4,3708	0,8545	1,2764	6,4543
<b>vmv</b>	30	13,3480	5,7998	4,4037	29,1158
<b>cpa</b>	60	8,1673	4,6734	3,5130	24,1553
<b>ttc</b>	63	100000	0,0000	100000	100000

Em todas as tabelas, é possível ver que o número de observações (N) é diferente. Tal deve-se ao facto de haver algumas variáveis omissas, como explicado anteriormente. Observa-se nas Tabelas 2,3 e 5 valores da variável vmp elevados que são explicados pelo atravessamento de um peão em corrida.

Além destas, foram consideradas ainda 3 variáveis (género, idade, movimento de cabeça) que são variáveis categóricas ou do tipo dummy, com os valores de 0 e 1. Para estas 3 variáveis, são agora apresentadas as respetivas frequências para todos os grupos de variáveis:

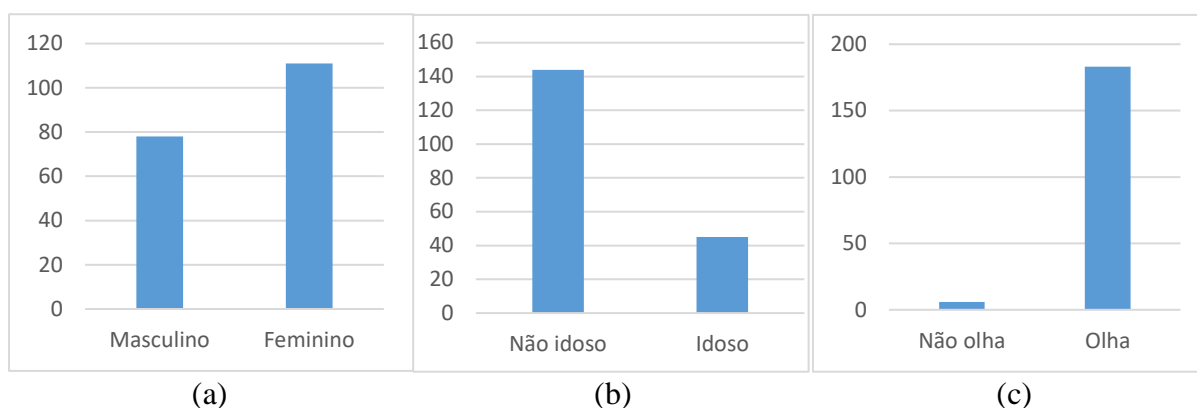


Figura 9. Frequência das variáveis dummy (grupo 1): (a) Género (b) Idade (c) Movimento de cabeça

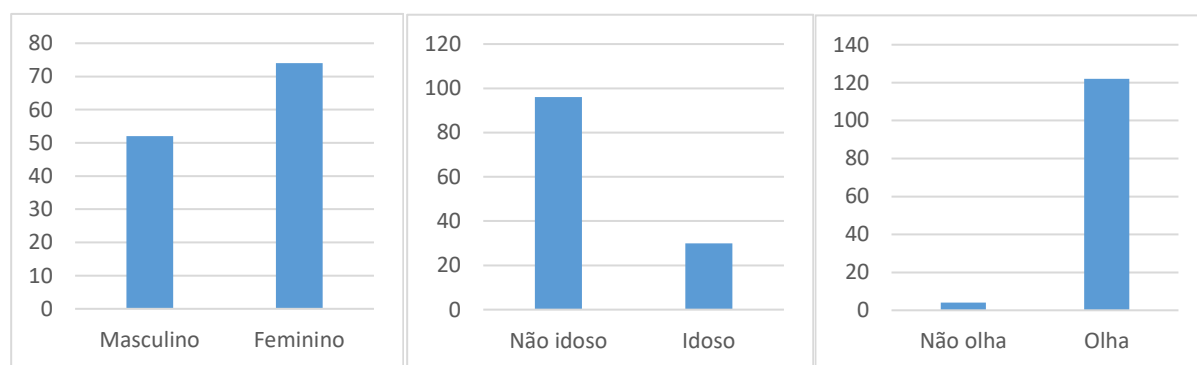


Figura 10. Frequência das variáveis dummy (grupo 2): (a) Género (b) Idade (c) Movimento de cabeça

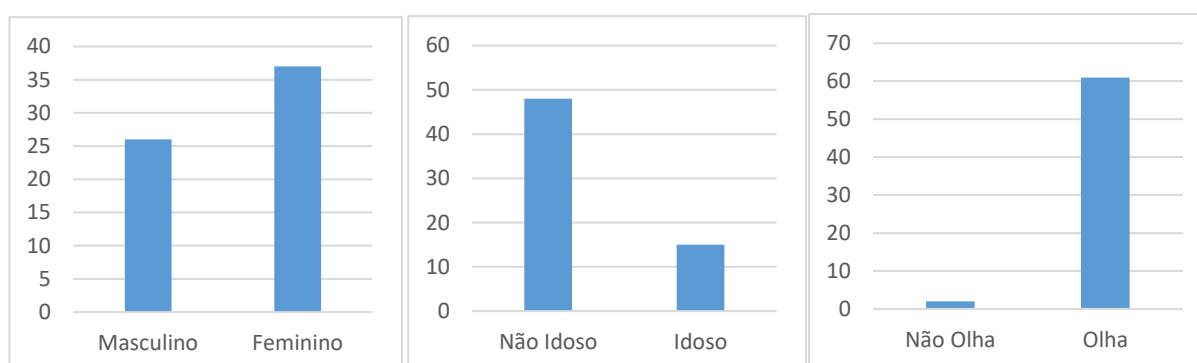


Figura 11. Frequência das variáveis dummy (grupo 3,4 e 5): (a) Género (b) Idade (c) Movimento de cabeça

Em todos os grupos de variáveis, verifica-se um número de não idosos bastante maior que o número de idosos o que poderá afetar alguns resultados na modelação, devido á dimensão da base de dados, numa fase mais avançada do presente trabalho. Para os grupos 3,4 e 5, cada um compreende o mesmo grupo de peões, variando apenas o momento do atravessamento, logo, as frequências das variáveis dummy, são todas iguais nesses três grupos.



## 4.2. Correlação entre variáveis

As matrizes que são apresentadas a seguir, determinam o coeficiente de correlação de Pearson. Este coeficiente assume valores entre -1 e 1, sendo que -1 significa que uma correlação negativa perfeita entre duas variáveis (se uma aumenta, a outra diminui sempre), e 1 significa uma correlação positiva perfeita entre duas variáveis. No caso de ser 0, significam que as duas variáveis não dependem linearmente uma da outra (Cohen et al., 2009).

Nas Tabelas 7,8,9,10 e 11, são apresentadas as matrizes de correlação entre as variáveis de cada grupo. Os valores que surgem de fundo azul, são referentes à variável dependente (vmp).

Tabela 7 - Matriz de correlação (Grupo 1)

\	ac	acv	viv	vip	vmp	vmv	cpa	ttc	genero	idade	mc
ac	1	0.046	-0.104	-0.600	0.090	-0.089	0.106	-0.298	0.045	0.007	0.005
acv	0.046	1	-0.135	-0.039	-0.018	-0.173	-0.380	0.384	-0.059	-0.051	0.003
viv	-0.104	-0.135	1	0.017	0.012	0.853	0.219	0.056	0.223	0.278	0.025
vip	-0.600	-0.039	0.017	1	0.248	<0.001	-0.118	0.162	-0.085	-0.020	<0.001
vmp	0.090	-0.018	0.012	0.248	1	-0.065	-0.144	0.120	-0.062	0.069	-0.102
vmv	-0.089	-0.173	0.853	<0.001	-0.065	1	0.557	-0.242	0.207	0.342	0.004
cpa	0.106	-0.380	0.219	-0.118	-0.144	0.557	1	-0.487	0.127	0.185	-0.020
ttc	-0.298	0.384	0.056	0.162	0.120	-0.242	-0.487	1	0.041	0.023	-0.037
genero	0.045	-0.059	0.223	-0.085	-0.062	0.207	0.127	0.041	1	0.241	0.216
idade	0.007	-0.051	0.278	-0.020	0.069	0.342	0.185	0.023	0.241	1	0.101
mc	0.005	0.003	0.025	<0.001	-0.102	0.004	-0.020	-0.037	0.216	0.101	1

Tabela 8 – Matriz de correlação (Grupo 2)

	ac	acv	viv	vip	vmp	vmv	cpa	ttc	genero	idade	mc
ac	1	0.230	-0.197	-0.454	0.277	-0.223	-0.151	0.131	0.107	0.042	-0.033
acv	0.230	1	-0.064	-0.051	0.051	-0.166	-0.225	0.104	-0.065	-0.017	-0.021
viv	-0.197	-0.064	1	-0.065	-0.166	0.840	0.404	-0.255	0.146	0.280	0.001
vip	-0.454	-0.051	-0.065	1	0.263	-0.034	-0.024	-0.129	-0.209	-0.038	-0.047
vmp	0.277	0.051	-0.166	0.263	1	-0.101	-0.142	0.065	-0.124	0.033	-0.195
vmv	-0.223	-0.166	0.840	-0.034	-0.101	1	0.586	-0.350	0.184	0.376	-0.011
cpa	-0.151	-0.225	0.404	-0.024	-0.142	0.586	1	-0.110	0.246	0.295	-0.092
ttc	0.131	0.104	-0.255	-0.129	0.065	-0.350	-0.110	1	0.123	-0.111	0.066
genero	0.107	-0.065	0.146	-0.209	-0.124	0.184	0.246	0.123	1	0.241	0.216
idade	0.042	-0.017	0.280	-0.038	0.033	0.376	0.295	-0.111	0.241	1	0.101
mc	-0.033	-0.021	0.001	-0.047	-0.195	-0.011	-0.092	0.066	0.216	0.101	1

Tabela 9 – Matriz de correlação (Grupo 3)

	<b>ac</b>	<b>acv</b>	<b>viv</b>	<b>vip</b>	<b>vmp</b>	<b>vmv</b>	<b>cpa</b>	<b>ttc</b>	<b>genero</b>	<b>idade</b>	<b>mc</b>
<b>ac</b>	1	0.223	-0.373	-0.417	0.337	-0.415	-0.108	0.559	0.116	0.012	-0.164
<b>acv</b>	0.223	1	-0.128	-0.056	-0.119	-0.244	-0.322	0.248	-0.229	-0.129	
<b>viv</b>	-0.373	-0.128	1	-0.001	-0.172	0.906	0.661	-0.462	0.042	0.323	
<b>vip</b>	-0.417	-0.056	-0.001	1	0.162	-0.009	-0.119	-0.388	-0.248	-0.061	0.025
<b>vmp</b>	0.337	-0.119	-0.172	0.162	1	-0.141	0.003	0.079	-0.088	0.042	-0.174
<b>vmv</b>	-0.415	-0.244	0.906	-0.009	-0.141	1	0.754	-0.294	0.101	0.4	
<b>cpa</b>	-0.108	-0.322	0.661	-0.119	0.003	0.754	1	-0.082	0.294	0.446	
<b>ttc</b>	0.559	0.248	-0.462	-0.388	0.079	-0.294	-0.082	1	0.085	-0.116	
<b>genero</b>	0.116	-0.229	0.042	-0.248	-0.088	0.101	0.294	0.085	1	0.242	0.216
<b>idade</b>	0.012	-0.129	0.323	-0.061	0.042	0.400	0.446	-0.116	0.242	1	0.101
<b>mc</b>	-0.164			0.025	-0.174				0.216	0.101	1

Tabela 10 – Matriz de correlação (Grupo 4)

	<b>ac</b>	<b>acv</b>	<b>viv</b>	<b>vip</b>	<b>vmp</b>	<b>vmv</b>	<b>cpa</b>	<b>ttc</b>	<b>genero</b>	<b>idade</b>	<b>mc</b>
<b>ac</b>	1	0.029	0.067	-0.576	0.042	0.029	-0.073	-0.147	0.121	0.124	0.245
<b>acv</b>	0.029	1	0.102	0.042	0.091	-0.079	0.026	0.080	0.133	0.107	0.010
<b>viv</b>	0.067	0.102	1	-0.178	-0.105	0.820	0.226	-0.209	0.231	0.276	-0.015
<b>vip</b>	-0.576	0.042	-0.178	1	0.569	-0.077	0.038	0.022	-0.164	-0.005	-0.170
<b>vmp</b>	0.042	0.091	-0.105	0.569	1	-0.086	-0.053	0.133	-0.175	0.027	-0.237
<b>vmv</b>	0.029	-0.079	0.820	-0.077	-0.086	1	0.512	-0.395	0.244	0.366	-0.016
<b>cpa</b>	-0.073	0.026	0.226	0.038	-0.053	0.512	1	-0.170	0.288	0.240	-0.179
<b>ttc</b>	-0.147	0.080	-0.209	0.022	0.133	-0.395	-0.170	1	0.160	-0.104	0.064
<b>genero</b>	0.121	0.133	0.231	-0.164	-0.175	0.244	0.288	0.160	1	0.242	0.216
<b>idade</b>	0.124	0.107	0.276	-0.005	0.027	0.366	0.240	-0.104	0.242	1	0.101
<b>mc</b>	0.245	0.010	-0.015	-0.170	-0.237	-0.016	-0.179	0.064	0.216	0.101	1

Tabela 11 – Matriz de correlação (Grupo 5)

	<b>ac</b>	<b>acv</b>	<b>viv</b>	<b>vip</b>	<b>vmp</b>	<b>vmv</b>	<b>cpa</b>	<b>ttc</b>	<b>genero</b>	<b>idade</b>	<b>mc</b>
<b>ac</b>	1	0.22	-0.018	-0.681	-0.035	-0.227	0.061		-0.009	-0.028	0.047
<b>acv</b>	0.22	1	0.189	-0.273	-0.292	0.420	-0.360		0.080	-0.033	0.224
<b>viv</b>	-0.018	0.189	1	0.080	0.335	0.866	0.063		0.332	0.272	0.057
<b>vip</b>	-0.681	-0.273	0.080	1	0.226	0.299	-0.107		0.071	0.008	0.062
<b>vmp</b>	-0.035	-0.292	0.335	0.226	1	0.097	0.014		0.094	0.165	0.129
<b>vmv</b>	-0.227	0.420	0.866	0.299	0.097	1	-0.247		0.304	0.039	-0.073
<b>cpa</b>	0.061	-0.360	0.063	-0.107	0.014	-0.247	1		0.304	0.039	-0.073
<b>ttc</b>								1			
<b>genero</b>	-0.009	0.080	0.332	0.071	0.094	0.304	0.304		1	0.242	0.216
<b>idade</b>	-0.028	-0.033	0.272	0.008	0.165	0.039	0.039		0.242	1	0.101
<b>mc</b>	0.047	0.224	0.057	0.062	0.129	-0.073	-0.073		0.216	0.101	1

De um modo geral não se verificam graus de correlação muito fortes. Em todas as tabelas, destacam-se as correlações verificadas entre a variável vmv e viv, com o valor a oscilar à volta de 0,8 indicando assim uma correlação forte, e ainda referente ao Grupo 3 na Tabela 8, uma correlação forte entre as variáveis cpa e vmv. Num grau mais moderado, verificam-se as correlações entre as variáveis vip e ac em todos os Grupos.

Nas Tabelas 9 e 11, não foi possível calcular a correlação para as variáveis mc e ttc respetivamente uma vez que ambas não têm variabilidade suficiente.

### 4.3. Testes de Normalidade

Os testes de normalidade são usados para determinar se um conjunto de dados é bem modelado pela distribuição normal ou não, e calcular a probabilidade de uma variável aleatória subjacente ao conjunto dos dados ser normalmente distribuída. Dependendo da forma como cada um interpreta as probabilidades, que pode variar, pode-se depreender que tais testes são uma forma de seleção de modelos.

Há vários tipos de teste de normalidade que podem ser usados. O presente estudo, serviu-se do teste de Shapiro-Wilk, que teste a hipótese nula de uma amostra  $x_1, \dots, x_n$  vir de uma população normalmente distribuída. O teste estatístico formula-se da seguinte forma:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (\text{Eq.15})$$

Onde,  $x_i$  é a  $i$ -ésima ordem estatística,  $\bar{x} = (x_1 + \dots + x_n)/n$  a média da amostra, e  $a_i$  constantes. Então, para um valor de  $\alpha=0,05$ , caso  $p < \alpha$  rejeita-se a hipótese da variável vmp ter uma distribuição normal. Se  $p > \alpha$  não se rejeita esta mesma hipótese (Shapiro & Wilk, 1965).

Então, para um valor de  $\alpha=0,05$ , caso  $p < \alpha$  rejeita-se a hipótese da variável vmp ter uma distribuição normal. Se  $p > \alpha$  não se rejeita esta mesma hipótese.

Na Tabela 12 são apresentados os resultados do teste de normalidade da variável dependente considerada.

Tabela 12 - Teste de normalidade da vmp

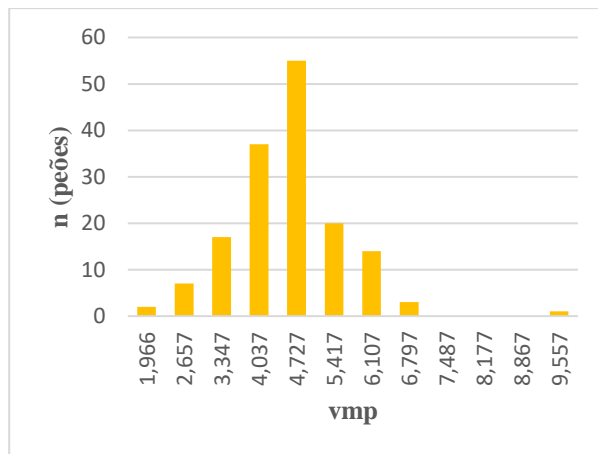
	<b>Grupo 1</b>	<b>Grupo 2</b>	<b>Grupo 3</b>	<b>Grupo 4</b>	<b>Grupo 5</b>
<b>W</b>	0.958	0.943	0.984	0.828	0.950
<b>p-value</b>	<0.001	<0.001	0.595	<0.001	0.013

No que se refere ao Grupo 1, o valor de  $p$  é igual a  $2,405e-05$ , e com base no que já foi mencionado anteriormente, pode-se afirmar que a hipótese de a variável vmp poder ter uma distribuição normal é rejeitada. O mesmo acontece para os Grupos 2, 4 e 5, onde o valor de  $p < 0,05$ , podendo afirmar-se, que para estes modelos a hipótese da variável vmp ter uma distribuição normal é rejeitada.

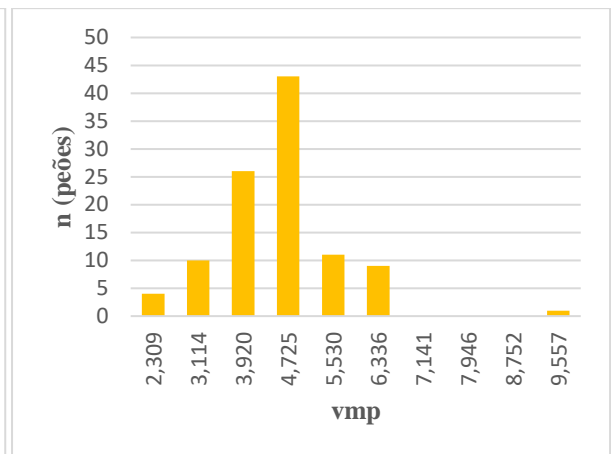
Apenas no Grupo 3, onde o valor de  $p$  é igual a  $0,5949$ , a hipótese da variável vmp ter uma distribuição normal não pode ser rejeitada.

Na Figura 12, são apresentados os histogramas da vmp para cada grupo de variáveis. Visualmente, os gráficos (a), (b) e (c) sugerem uma distribuição normal dos dados. Nos primeiros dois há um valor extremo à direita que se deve, como já foi referido, ao atravessamento de um peão em corrida.

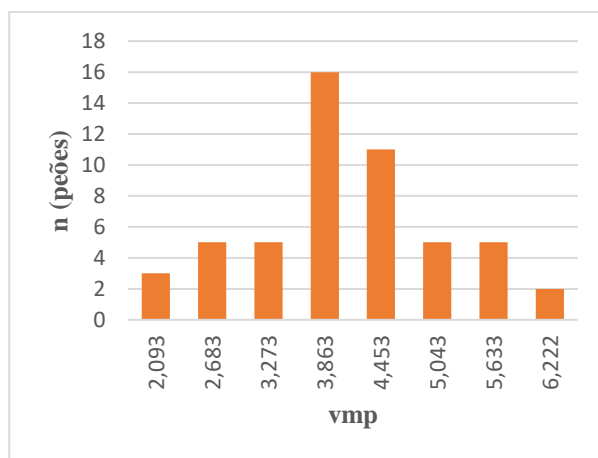
Nos Grupos 1,2 pode-se observar que a maioria dos peões tem uma vmp de aproximadamente  $4,7$  km/h, e no Grupo 4 cerca de  $4,5$  km/h. No Grupo 3, a maioria tem uma vmp de  $3,8$  km/h e no Grupo 5 de  $5,16$  km/h. Ou seja, antes do atravessamento as velocidades médias do peão são inferiores às depois do atravessamento.



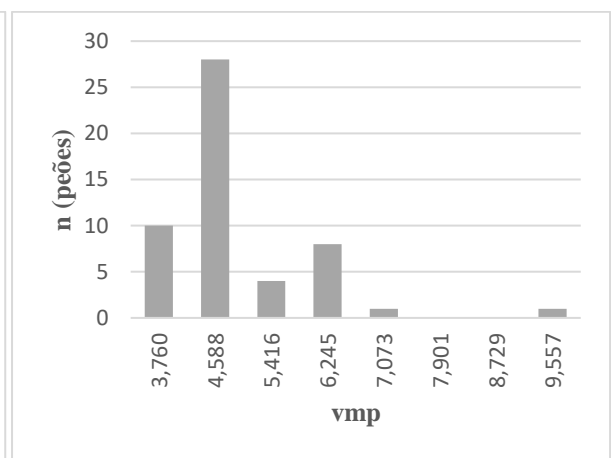
(a)



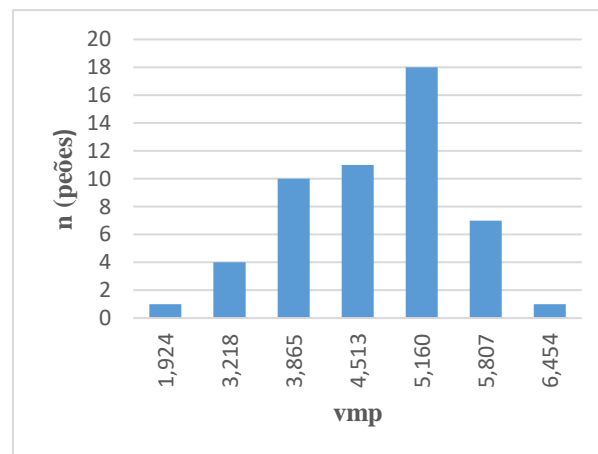
(b)



(c)



(d)



(e)

Figura 12.(a) Histograma da vmp (Grupo 1); (b) Histograma da vmp (Grupo 2); (c) Histograma da vmp (Grupo 3); (d) Histograma da vmp (Grupo 4); (e) Histograma da vmp (Grupo 5)

## 4.5. Resultados dos modelos

Nesta etapa, o objetivo principal passa por observar os efeitos das variáveis no comportamento dos peões em atravessamentos. Mais concretamente, identificar os fatores que têm impacto na velocidade média do peão, a partir da aplicação da técnica de modelação mais adequada.

A técnica das equações de estimação generalizadas (EEG) foi a abordagem escolhida uma vez que se está a tratar o atravessamento em 3 momentos de forma agregada e desagregada. Refere-se que são uma extensão dos modelos lineares generalizados, detalhados na revisão bibliográfica, pelo que pode ser aplicada aos grupos de dados desagregados (grupo 3 a 5).

### 4.5.1. Resultados modelo 1

Para o modelo 1 ou agregado, foram considerando os 3 momentos de atravessamento juntos. Num total de 189 observações lidas, apenas 162 foram usadas, ou seja, 27 observações ficaram de fora deste modelo. A variável dependente definida (vmp), considera a componente de erro uma distribuição binomial negativa. A escolha das variáveis que irão incorporar o modelo estatística é um processo iterativo, onde em primeira instância são consideradas todas as variáveis descritas no subcapítulo da definição das variáveis. As variáveis que não se revelam estatisticamente significativas vão sendo retiradas até se encontrar o modelo mais ajustado. Na Tabela 13 é apresentado o resultado deste primeiro modelo:

Tabela 13 - Resultados GEE modelo 1

<b>Parâmetros</b>	<b>Estimativa</b>	<b>Desvio padrão</b>	<b>Pr &gt;  Z </b>
<b>Intercept</b>	1.6375	0.0334	<.0001
<b>cpa</b>	-0.0055	0.0022	0.0131
<b>idade</b>	0.0628	0.0357	0.0788
<b>mc</b>	-0.1517	0.0357	<.0001

Para gerar estes resultados, os níveis de significância considerados foram de 5%, admitindo-se um máximo de 10%. Como se pode observar, o resultado contempla apenas três variáveis estatisticamente significativas.

As variáveis  $cpa$  e  $mc$  mostram um impacto negativo na velocidade média do peão ( $vmp$ ), ou seja, a sua relação é inversamente proporcional.

As hipóteses assumidas para esclarecer esta relação para o caso do  $cpa$  são as seguintes:

- Considerando uma situação específica em que o peão está a realizar um atravessamento e vê um veículo aproximar-se (a distância entre os objetos é menor), acelera o passo alcançando uma velocidade maior com o propósito de evitar um possível conflito entre os dois.
- Por outro lado, quando o  $cpa$  é maior, a velocidade média do peão é menor, o que pode significar que o peão tem uma velocidade média mais baixa ao ver um veículo mais distante, sentindo que não está em perigo e consequentemente sem necessidade de apressar o seu passo para efetuar o atravessamento e evitar um conflito.

No caso da variável  $mc$ , os resultados podem ser interpretados da seguinte forma:

- Caso o valor dela seja 1 (o peão olha), a velocidade média do peão é inferior. Então, quando o peão vai fazer um atravessamento e olha, a sua velocidade média é inferior, o que pode indiciar prudência da sua parte.
- No caso de ser 0 (o peão não olha) e a sua velocidade média é superior. Tal pode acontecer porque o peão antecipadamente já tinha percecionado o espaço à sua volta e se tenha apercebido que não haviam veículos por perto naquele momento. Por outro lado pode ser percebido como um sinal de negligência por parte do peão, onde não olha e como tal a sua velocidade é superior com o objetivo de terminar o atravessamento o mais rápido possível.

Já a variável  $idade$ , sugere um resultado mais questionável, ou seja, não era expectável de acordo com a literatura. Indiciando que a relação entre a idade e a variável dependente é diretamente proporcional, ou seja, revela um impacto positivo na relação entre a idade do peão e a sua velocidade média. A justificação deste resultado, prende-se sobretudo com a dimensão da base de dados. Uma vez que amostra é curta, resultados tendenciosos como este podem aparecer. O facto da variável  $idade$  compreender apenas 2 grupos (idosos e não idosos) também pode ter influência, além de que o número de idosos é mais baixo que o número de não idosos que contemplam crianças e adultos no mesmo grupo.

### 4.5.2. Resultados Modelo 2

No modelo 2, que junta os momentos “antes” e “durante”, o número de observações lidas foi de 126, sendo que 100 foram utilizadas. A variável dependente manteve-se a vmp, considerando a componente de erro uma distribuição binomial negativa. O procedimento seguido para o modelo 2 é o mesmo que foi utilizado no modelo 1, eliminando as variáveis que não se mostravam significativas. O resultado pode ser visto na Tabela 14:

Tabela 14 - Resultados GEE modelo 2

<b>Parâmetros</b>	<b>Estimativa</b>	<b>Desvio padrão</b>	<b>Pr &gt;  Z </b>
<b>Intercept</b>	1.9235	0.0881	<.0001
<b>vmv</b>	0.0099	0.0054	0.0679
<b>viv</b>	-0.0099	0.0037	0.0076
<b>cpa</b>	-0.0092	0.0045	0.0392
<b>mc</b>	-0.3895	0.0397	<.0001

Os níveis de significância mantiveram-se nos 5%, admitindo-se um máximo de 10%. Neste modelo, há mais uma variável significativa que no anterior, ou seja, são 4 no total.

As variáveis viv, cpa e mc mostram um impacto negativo na velocidade média do peão. Sendo a vmv, a única variável diretamente proporcional à variável dependente.

A variável vmv, contribui positivamente para a velocidade média do peão. Uma forma de analisar esta relação, reside na ideia de que para um valor da vmv superior, o peão fica mais alerta e consequentemente tem a sua velocidade média também superior. No mesmo raciocínio, vmv inferiores irão proporcionar uma sensação de maior segurança no peão, traduzindo-se assim em vmp menores.

Quanto à variável viv, os resultados revelaram-se duvidosos, uma vez que para esta variável a relação mostra um impacto negativo relativamente à variável dependente, ao contrário da vmv. Em trabalhos futuros é aconselhada uma atenção especial para esta situação.

A justificação encontrada para este resultado está relacionada com a técnica estatística utilizada, a GEE. Apesar de este modelo contemplar mais que um momento de observação, a GEE considera os dois espaços temporais. Isto é, a variável viv no momento “antes”, é relativa à velocidade 2 segundos antes do peão iniciar o atravessamento, e no momento “durante” é



relativa ao momento exato do início do atravessamento do mesmo. Nesta situação, o modelo considera os dois momentos separadamente, logo, pode-se concluir que:

- Para 2 segundos antes do peão iniciar o atravessamento e no momento exato em que ele inicia a travessia, a viv tem um impacto negativo na velocidade média do peão. Uma explicação para esta situação reside na ideia de que, nestes dois momentos em que o peão se depara com o veículo, ele tem uma velocidade maior para que possa dar início à travessia antes que o veículo chegue primeiro, uma vez que sente que pode alcançar a passadeira antes do veículo e assim ter a prioridade no atravessamento.
- Ao contrário, quando a viv é maior, o peão irá ter uma vmp inferior sentindo que não consegue iniciar a travessia em segurança.

Para evitar este resultado mais duvidoso, considerou-se a hipótese de retirar a variável vmv ou viv do modelo. Mas ao fazê-lo, o modelo tornou-se instável e por isso mantiveram-se ambas as variáveis. Quanto às variáveis cpa e mc, o resultado demonstra um comportamento igual ao obtido no modelo 1, onde ambas mostram um impacto negativo na vmp. Portanto, a interpretação destes resultados segue a mesma lógica encontrada para o modelo 1.

#### **4.5.3. Resultados dos Modelo 3 a 5**

A análise dos grupos de variáveis 3,4 e 5 (desagregados), mostrou que não haviam variáveis significativas. Desta forma, é uma abordagem que não serve este caso de estudo específico.

#### **4.6. Análise de resíduos**

Uma análise de resíduos, é uma abordagem comumente utilizada em modelos de regressão que permite avaliar a qualidade do ajuste do modelo. O resíduo podem ser definido como a diferença entre o valor observado e o valor estimado da variável dependente(Costa, 2013).

Nas Figuras 13 e 14 são apresentados os resíduos para ambos os modelos para verificar então o ajuste dos modelos.

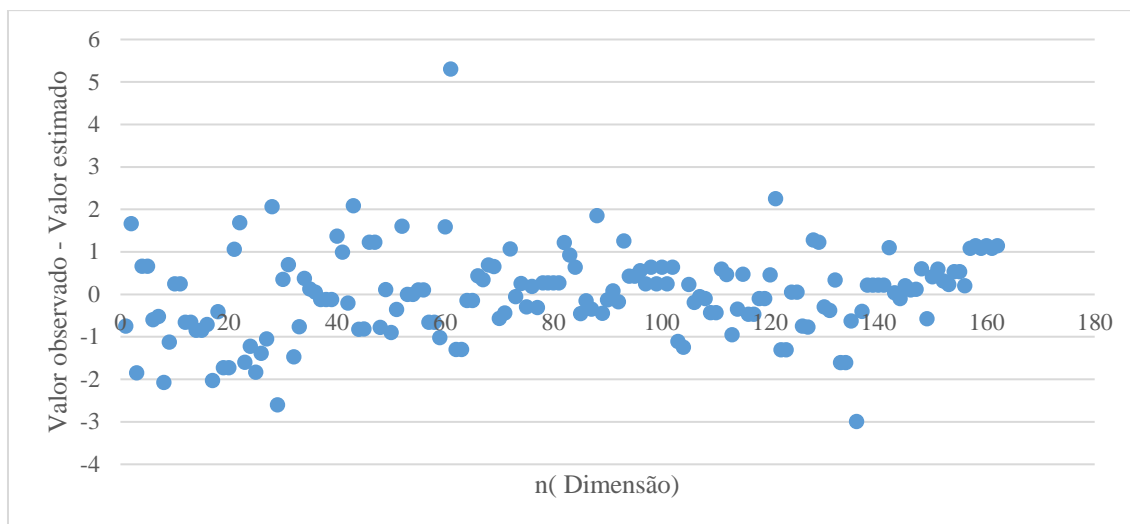


Figura 13. Análise de resíduos Modelo 1

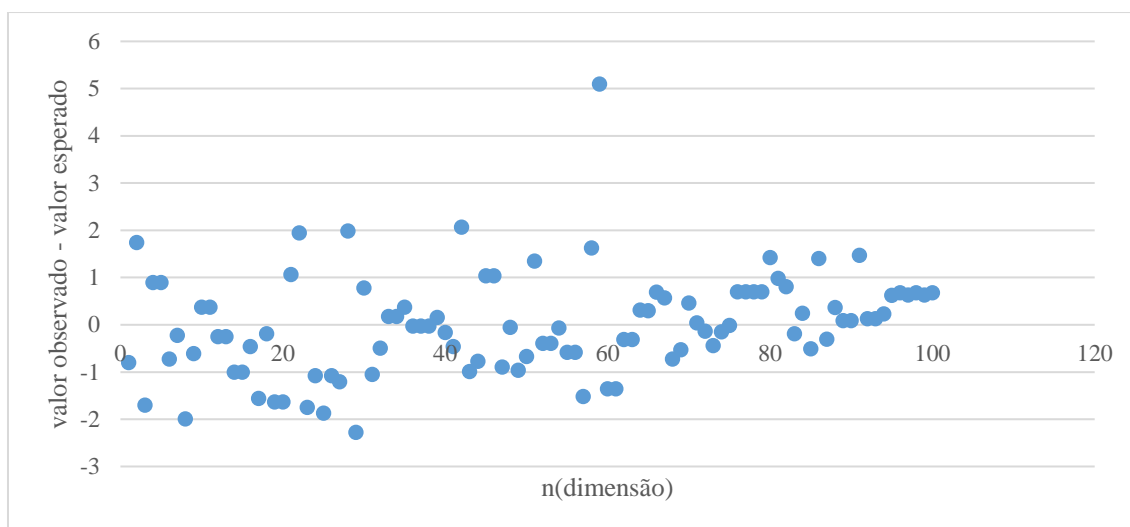


Figura 14. Análise de resíduos Modelo 2

Como pretendido, os resíduos variam em torno do zero. Um resíduo que varia em torno do zero, sem grande oscilação, indicia que o modelo não é tendencioso, ou seja, demonstra resultados fidedignos.

Nas Figuras 13 e 14, observa-se uma distribuição de valores idêntica com uma variação entre -2 e 2. O cálculo da raiz média quadrática, que resultou num valor de 1,01 e 1,06 para os modelos 1 e 2 respetivamente, é indicativo que as diferenças são reduzidas e ambos os modelos sugerem um ajuste idêntico.

## 5. CONCLUSÕES E DESENVOLVIMENTOS FUTUROS

### 5.1. Conclusões

Neste último capítulo será apresentada uma síntese dos principais temas abordados e respectivas conclusões. Serão expostas as limitações encontradas assim como propostas de melhoria e alterações metodológicas que possam tornar este tipo de análises mais eficazes.

A primeira etapa num trabalho deste tipo começa com uma rigorosa revisão bibliográfica, onde se estudam os modelos estatísticos mais utilizados na literatura. Constata-se que existe uma vasta gama de modelos estatísticos, com diferentes aplicabilidades e no caso específico de modelos de comportamento dos peões há ainda uma grande margem de progressão. É portanto fundamental dominar os conceitos fundamentais da modelação para que se obtenham resultados que permitam implementar medidas eficazes no combate à sinistralidade rodoviária.

A divisão dos vários tipos de modelos não é direta e pode variar de autor para autor. Neste caso, dividiram-se em modelos de comportamento dos peões e modelos de gravidade e frequência dos acidentes. A escolha de um modelo ajustado depende de vários aspetos, tais como a tipologia e natureza da base de dados, revelando-se assim uma tarefa relativamente difícil.

Um outro grande obstáculo na modelação prende-se com a construção da base de dados por diversas razões, que podem estar relacionadas com a imprevisibilidade da natureza humana e ainda limitações de logística. Neste trabalho esta etapa foi realizada com o auxílio do programa *Traffic Intelligence* desenvolvido pelo canadiano *Nicolas Saunier*, que se revelou muito útil uma vez que tem a capacidade de identificar objetos em movimento e classifica-los em veículos ou peões. Define trajetórias para ambos os objetos num eixo bidimensional e consegue medir velocidades. Assim, com base na literatura foi definido um conjunto de variáveis que caracterizem tanto os veículos como os peões. Foram elaborados 5 grupos de variáveis que consideram 3 momentos de observação previamente definidos: antes, durante e depois do atravessamento.

Uma vez completa a base de dados, fez-se um estudo estatístico da base de dados através de técnicas de estatística simples e descritiva aos 5 grupos de variáveis. Também se fizeram testes de normalidade e matrizes de correlação, com o objetivo de descrever e encontrar padrões nos dados recolhidos. Esta fase é muito importante uma vez que para a escolha da técnica estatística a ser aplicada, é necessário conhecer as características dos dados para que se possam obter

resultados conclusivos. Através do programa estatístico SAS, partiu-se para a modelação dos dados com a aplicação da técnica equações de estimativas generalizadas. Para os grupos de variáveis 3,4 e 5, a modelação mostrou-se inconclusiva. Para os grupos 1 e 2, a modelação foi feita com a variável dependente definida como a velocidade média do peão, que juntamente com um conjunto de variáveis explicativas forneceram resultados conclusivos.

É importante referir que o objetivo principal deste estudo consiste na identificação dos fatores que intervêm no comportamento dos peões no momento de um atravessamento. E partindo deste pressuposto, os resultados demonstraram que há um conjunto de variáveis que têm impacto na variável dependente.

Os resultados revelaram que o facto de um peão olhar ou não no momento de um atravessamento, a idade desse mesmo peão, ou a velocidade do veículo têm influência na velocidade média do peão. Ou seja, o comportamento que os peões adotam quando fazem um atravessamento numa passadeira é afetado por um conjunto de outros fatores que podem eventualmente ser controlados através de medidas de segurança rodoviária.

Quanto à variável idade, devido à curta dimensão da base de dados ela mostrou pouca variância e como tal os resultados não foram os esperados. As variáveis relativas à velocidade dos veículos conduziram a resultados aparentemente contraditórios, mas essenciais para a estabilidade do modelo. A utilização de uma base de dados de maior dimensão permitirá aferir os resultados obtidos.

Outros tipos de abordagens podem ser feitos em estudos desta génese. Para este estudo foi adotada uma análise de natureza quantitativa embora se possa abordar o comportamento dos peões de forma qualitativa. Tal deve-se ao programa que aqui foi utilizado que origina dados quantitativos e como tal trabalhou-se neste sentido.

## **5.2. Limitações e desenvolvimentos futuros**

Ao longo do trabalho surgiram várias limitações, principalmente relacionadas com a construção da base de dados. A extração de dados feita a partir de gravações de vídeo de ambos os locais, mostrou que a posição da câmara é fundamental para que o programa consiga mais eficazmente identificar os objetos e definir as suas trajetórias. O programa opera a partir do movimento dos objetos e é muito sensível a qualquer tipo de movimento. É por isso importante ter em atenção esta sensibilidade assim como as condições climatéricas para evitar zonas com

muita vegetação suscetíveis à ação do vento. Convém também evitar zonas com planos de reflexão devido à sensibilidade do programa.

A proximidade da câmara ao local de gravação é também relevante uma vez que caso esteja muito distante, a identificação dos objetos torna-se menos precisa. Nem sempre é fácil conseguir uma localização perfeita pelo facto de o local ideal ser muitas vezes difícil de alcançar. Em estudos futuros que se sirvam deste tipo de análise recomenda-se a escolha de um local que permita um posicionamento tendo em conta as fatores anteriormente referidos. Na literatura podem ser encontrados artigos que detalham com mais rigor de que forma se pode ter um posicionamento bom. Há que ter em conta que o programa *Traffic Intelligence* se encontra em desenvolvimento e é de livre acesso, para que possa ser melhorado.

O bom funcionamento do programa depende muito de uma boa calibração dos parâmetros de deteção dos objetos. Além de não haver muita informação sobre como obter uma calibração, é difícil de a conseguir.

Depois de modelados os dados, e como foi referido nesse capítulo, há um resultado que se mostrou um pouco duvidoso. Algumas explicações foram encontradas para justificar esse mesmo resultado, mas o que é importante reter, é que a definição das variáveis a incorporar no modelo estatístico deve ser cuidadosamente elaborada.

Quanto à dimensão da base de dados, o facto de não ser muito extensa pode gerar menos eficiência nos resultados, no entanto, como este trabalho é de natureza exploratória e com poucas similitudes na literatura existentes, pode ser visto como um bom ponto de partida para análises desta génese.



## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research*, 34(5), 597–603. <https://doi.org/10.1016/j.jsr.2003.05.009>
- Abdel-aty, M., & Abdalla, F. (2004). Linking Roadway Geometrics and Real-Time Traffic Characteristics to Model Daytime Freeway Crashes Generalized Estimating Equations for Correlated Data. *Transportation Research Record: Journal of the Transportation Research Board*, 1897, 106–115. <https://doi.org/10.3141/1897-14>
- Aguero-Valverde, J., & Jovanis, P. P. (2009). Bayesian Multivariate Poisson Lognormal Models for Crash Severity Modeling and Site Ranking. *Transportation Research Record: Journal of the Transportation Research Board*, 2136(1), 82–91. <https://doi.org/10.3141/2136-10>
- Al-Madani, H., & Al-Janahi, A. (2006). Personal exposure risk factors in pedestrian accidents in Bahrain. *Safety Science*, 44(4), 335–347. <https://doi.org/10.1016/j.ssci.2005.10.009>
- Anastasopoulos, P. C., & Mannering, F. L. (2009). A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis and Prevention*, 41(1), 153–159. <https://doi.org/10.1016/j.aap.2008.10.005>
- Anastasopoulos, P. C., & Mannering, F. L. (2011). An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data. *Accident Analysis and Prevention*, 43(3), 1140–1147. <https://doi.org/10.1016/j.aap.2010.12.024>
- Antonini, G., Bierlaire, M., & Weber, M. (2006). Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, 40(8), 667–687. <https://doi.org/10.1016/j.trb.2005.09.006>
- Autoridade Nacional Segurança Rodoviária. (2015). *Sinistralidade Rodoviária-Ano de 2015*. Retrieved from [http://www.ansr.pt/Estatisticas/RelatoriosDeSinistralidade/Documents/2015/RELATÓRIO ANUAL - VÍTIMAS A 30 DIAS/Rel2015\\_anual30dias.pdf](http://www.ansr.pt/Estatisticas/RelatoriosDeSinistralidade/Documents/2015/RELATÓRIO ANUAL - VÍTIMAS A 30 DIAS/Rel2015_anual30dias.pdf)
- Ben-akiva, M., & Bierlaire, M. (2003). 2 Discrete Choice Models With Applications To Departure Time and Route Choice, 7–38.
- Bernhoft, I. M., & Carstensen, G. (2008). Preferences and behaviour of pedestrians and cyclists by age and gender. *Transportation Research Part F: Traffic Psychology and Behaviour*, 11(2), 83–95. <https://doi.org/10.1016/j.trf.2007.08.004>
- Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics*, 46(3), 347–364. [https://doi.org/10.1016/0304-4076\(90\)90014-K](https://doi.org/10.1016/0304-4076(90)90014-K)
- Cao, X., Mokhtarian, P. L., & Handy, S. L. (2007). Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations modeling approach. *Transportation*, 34(5), 535–556. <https://doi.org/10.1007/s11116-007-9132-x>
- Cohen, I., Huang, Y., Chen, J., & Benesty, J. (2009). Noise Reduction in Speech Processing, 2, 3–6. <https://doi.org/10.1007/978-3-642-00296-0>
- Costa, J. O. da. (2013). Desenvolvimento dum modelo de previsão de acidentes, 296.

- Delen, D., Sharda, R., & Bessonov, M. (2006). Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accident Analysis and Prevention*, 38(3), 434–444. <https://doi.org/10.1016/j.aap.2005.06.024>
- Duncan, C. S., Khattak, A. J., & Council, F. M. (1998). Applying the ordered probit model to injury severity in truck – passenger car rear-end collisions. *Transportation Research Record: Journal of the Transportation Research Board*, 1635(98), 63–71. <https://doi.org/10.3141/1635-09>
- El-Basyouny, K., Sayed, T., & Trb. (2006). Comparison of two negative binomial regression techniques in developing accident prediction models. *Statistical Methods and Crash Prediction Modeling*, (January 2006), 9–16. <https://doi.org/10.3141/1950-02>
- Elliot, M., & Tranmer, M. (n.d.). Binary Logistic Regression.
- EUROSTAT. (2015). Key trends (2015 monitoring). Retrieved February 13, 2017, from <http://ec.europa.eu/eurostat/web/sdi/indicators/sustainable-transport>
- Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression Analyses of Counts and Rates: Poisson, Overdispersed Poisson, and Negative Binomial Models. *Psychological Bulletin*, 118(3), 392–404. <https://doi.org/10.1037/0033-2909.118.3.392>
- Hanley, J. A., Negassa, A., Edwardes, M. D. deB., & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: An orientation. *American Journal of Epidemiology*, 157(4), 364–375. <https://doi.org/10.1093/aje/kwf215>
- Harney, D. (2002). Pedestrian modelling: current methods and future predictions. *Road and Transport Research*, 11(4), 38–48. Retrieved from <http://arrbknowledge.com>
- Hensher, D. a, & Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation*, 30, 133–176. <https://doi.org/10.1023/A:1022558715350>
- Jackson, S., Miranda-Moreno, L., St-Aubin, P., & Saunier, N. (2013). Flexible, Mobile Video Camera System and Open Source Video Analysis Software for Road Safety and Behavioral Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2365, 90–98. <https://doi.org/10.3141/2365-12>
- Jones, B., Janssen, L., & Mannering, F. (1991). Analysis of the frequency and duration of freeway accidents in Seattle. *Accident Analysis and Prevention*, 23(4), 239–255. [https://doi.org/10.1016/0001-4575\(91\)90003-N](https://doi.org/10.1016/0001-4575(91)90003-N)
- Jones, S., & Hensher, D. A. (2007). Evaluating the behavioural performance of alternative logit models: An application to corporate takeovers research. *Journal of Business Finance and Accounting*, 34(7–8), 1193–1220. <https://doi.org/10.1111/j.1468-5957.2007.02049.x>
- Jovanis, P. P., & Chang, H.-L. (1986). Modeling the relationship of accidents to miles traveled. *Transportation Research Board*, 45–51.
- Kim, D. G., Lee, Y., Washington, S., & Choi, K. (2007). Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis and Prevention*, 39(1), 125–134. <https://doi.org/10.1016/j.aap.2006.06.011>
- Kim, J. K., Ulfarsson, G. F., Shankar, V. N., & Mannering, F. L. (2010). A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accident*



- Analysis and Prevention*, 42(6), 1751–1758. <https://doi.org/10.1016/j.aap.2010.04.016>
- Kline, R. B. (2010). *Principles and practice of structural equation modeling. Structural Equation Modeling* (Vol. 156). <https://doi.org/10.1038/156278a0>
- Kockelman, K. M., & Kweon, Y. J. (2002). Driver injury severity: An application of ordered probit models. *Accident Analysis and Prevention*, 34(3), 313–321. [https://doi.org/10.1016/S0001-4575\(01\)00028-8](https://doi.org/10.1016/S0001-4575(01)00028-8)
- Kumara, S. S. P., & Chin, H. C. (2003). Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. *Traffic Injury Prevention*, 4(1), 53–7. <https://doi.org/10.1080/15389580309852>
- Lee, C., & Abdel-Aty, M. (2005). Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida. *Accident Analysis and Prevention*, 37(4), 775–786. <https://doi.org/10.1016/j.aap.2005.03.019>
- Lee, J., & Mannering, F. (2002). Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis. *Accident Analysis and Prevention*, 34(2), 149–161. [https://doi.org/10.1016/S0001-4575\(01\)00009-4](https://doi.org/10.1016/S0001-4575(01)00009-4)
- Lenguerrand, E., Martin, J. L., & Laumon, B. (2006). Modelling the hierarchical structure of road crash data - Application to severity analysis. *Accident Analysis and Prevention*, 38(1), 43–53. <https://doi.org/10.1016/j.aap.2005.06.021>
- Lerman, S. R., & Gonzalez, S. L. (1980). Poisson Regression Analysis Under Alternate Sampling Strategies. *Transportation Science*, 14(4), 346–364. <https://doi.org/10.1287/trsc.14.4.346>
- Liao, T. F. (1994). *Interpreting Probability Models Logit , Probit , and Other Generalized Linear Models*. Sage Publications, Inc.
- Lord, D. (2006). Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis and Prevention*, 38(4), 751–766. <https://doi.org/10.1016/j.aap.2006.02.001>
- Lord, D., Geedipally, S. R., & Guikema, S. D. (2010). Extension of the application of conway-maxwell-poisson models: Analyzing traffic crash data exhibiting underdispersion. *Risk Analysis*, 30(8), 1268–1276. <https://doi.org/10.1111/j.1539-6924.2010.01417.x>
- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291–305. <https://doi.org/10.1016/j.tra.2010.02.001>
- Lord, D., & Persaud, B. (2000). Accident Prediction Models With and Without Trend: Application of the Generalized Estimating Equations Procedure. *Transportation Research Record: Journal of the Transportation Research Board*, 1717(0), 102–108. <https://doi.org/10.3141/1717-13>
- Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. *Accident Analysis and Prevention*, 37(1), 35–46. <https://doi.org/10.1016/j.aap.2004.02.004>
- Ma, J., Kockelman, K. M., & Damien, P. (2008). A multivariate Poisson-lognormal

- regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis and Prevention*, 40(3), 964–975.  
<https://doi.org/10.1016/j.aap.2007.11.002>
- Malyshkina, N. V. (2008). Markov switching models: an application to roadway safety, (December), 122. Retrieved from <http://arxiv.org/abs/0808.1448>
- Malyshkina, N. V., & Mannering, F. (2009). Effect of Increases in Speed Limits on Severities of Injuries in Accidents. *Transportation Research Record: Journal of the Transportation Research Board*, 2083(1), 122–127. <https://doi.org/10.3141/2083-14>
- Malyshkina, N. V., & Mannering, F. L. (2010). Zero-state Markov switching count-data models: An empirical assessment. *Accident Analysis and Prevention*, 42(1), 122–130. <https://doi.org/10.1016/j.aap.2009.07.012>
- Matsui, Y., Hitosugi, M., Doi, T., Oikawa, S., Takahashi, K., & Ando, K. (2013). Features of Pedestrian Behavior in Car-to-Pedestrian Contact Situations in Near-Miss Incidents in Japan. *Traffic Injury Prevention*, 14(SUPPL1), 37–41.  
<https://doi.org/10.1080/15389588.2013.796372>
- Miaou, S. P., & Lum, H. (1993). Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis and Prevention*, 25(6), 689–709.  
[https://doi.org/10.1016/0001-4575\(93\)90034-T](https://doi.org/10.1016/0001-4575(93)90034-T)
- Morse, P., Student, M., Building, M. E., West, S. S., & Ha, C. (n.d.). Transferability Study of Video Tracking Optimization for Traffic Data Collection and Analysis, (514).
- Oh, J., Washington, S. P., & Nam, D. (2006). Accident prediction model for railway-highway interfaces. *Accident Analysis and Prevention*, 38(2), 346–356.  
<https://doi.org/10.1016/j.aap.2005.10.004>
- Olszewski, P., Szagała, P., Wolański, M. I., & Zielińska, A. (2015). Pedestrian fatality risk in accidents at unsignalized zebra crosswalks in Poland. *Accident Analysis and Prevention*, 84, 83–91. <https://doi.org/10.1016/j.aap.2015.08.008>
- Papadimitriou, E. (2012). Theory and models of pedestrian crossing behaviour along urban trips. *Transportation Research Part F: Traffic Psychology and Behaviour*, 15(1), 75–94.  
<https://doi.org/10.1016/j.trf.2011.11.007>
- Papadimitriou, E., Yannis, G., & Golias, J. (2009). A critical assessment of pedestrian behaviour models. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(3), 242–255. <https://doi.org/10.1016/j.trf.2008.12.004>
- Park, E. S., & Lord, D. (2008). Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. *Transportation Research Record*, 2019(1), 1–6.  
<https://doi.org/10.3141/2019-01>
- Perryman, M. R. (n.d.). *Econometric Modeling in Economic Education Research*.
- Poch, M., & Fred, M. (1996). NEGATIVE BINOMIAL ANALYSIS OF INTERSECTION-ACCIDENT FREQUENCIES, (April), 105–113.
- Quddus, M. A., Noland, R. B., & Chin, H. C. (2002). An analysis of motorcycle injury and vehicle damage severity using ordered probit models. *Journal of Safety Research*, 33(4), 445–462. [https://doi.org/10.1016/S0022-4375\(02\)00051-8](https://doi.org/10.1016/S0022-4375(02)00051-8)
- Rankavat, S., & Tiwari, G. (2016). Pedestrians risk perception of traffic crash and built

- environment features - Delhi, India. *Safety Science*, 87, 1–7.  
<https://doi.org/10.1016/j.ssci.2016.03.009>
- Robin, T., Antonini, G., Bierlaire, M., & Cruz, J. (2009). Specification, estimation and validation of a pedestrian walking behavior model. *Transportation Research Part B: Methodological*, 43(1), 36–56. <https://doi.org/10.1016/j.trb.2008.06.010>
- Savolainen, P., & Mannering, F. (2007). Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. *Accident Analysis and Prevention*, 39(5), 955–963. <https://doi.org/10.1016/j.aap.2006.12.016>
- Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis and Prevention*, 43(5), 1666–1676.  
<https://doi.org/10.1016/j.aap.2011.03.025>
- Schneider, R. J., Ryznar, R. M., & Khattak, A. J. (2004). An accident waiting to happen: A spatial approach to proactive pedestrian planning. *Accident Analysis and Prevention*, 36(2), 193–211. [https://doi.org/10.1016/S0001-4575\(02\)00149-5](https://doi.org/10.1016/S0001-4575(02)00149-5)
- Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.2307/1267427>
- Tay, R., Choi, J., Kattan, L., & Khan, A. (2011). A Multinomial Logit Model of Pedestrian-Vehicle Crash Severity. *International Journal of Sustainable Transportation*, 5(4), 233–249. <https://doi.org/10.1080/15568318.2010.497547>
- V. Shankar, Milton, J., & F. Mannering. (1997). Modeling Accident Frequencies as Zero-Altered Probability Processes : An Empirical Inquiry, 29(206), 829–837.  
[https://doi.org/10.1016/S0001-4575\(97\)00052-3](https://doi.org/10.1016/S0001-4575(97)00052-3)
- Van Acker, V., Witlox, F., & Van Wee, B. (2007). The Effects of the Land Use System on Travel Behavior: A Structural Equation Modeling Approach. *Transportation Planning and Technology*, 30(4), 331–353. <https://doi.org/10.1080/03081060701461675>
- Walker, J. (2001). Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables. *Environmental Engineering*, 208.  
<https://doi.org/http://transp-or.epfl.ch/courses/dca2012/WalkerPhD.pdf>
- Washington, S. P., Karlaftis, M. G., & Mannering, F. L. (2011). *Statistical, Econometric Methods: Transportation Data Analysis*.
- Yannis, G., Golias, J., & Papadimitriou, E. (2007). Modeling Crossing Behavior and Accident Risk of Pedestrians. *Journal of Transportation Engineering*, 133(11), 634–644.  
[https://doi.org/10.1061/\(ASCE\)0733-947X\(2007\)133:11\(634\)](https://doi.org/10.1061/(ASCE)0733-947X(2007)133:11(634))
- Ye, F., & Lord, D. (2014). Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. *Analytic Methods in Accident Research*, 1, 72–85. <https://doi.org/10.1016/j.amar.2013.03.001>

